OCT 3 1 1945

PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

Vol. XXXVI OCTOBER, 1945 No. 7

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

On the Permissible Coarseness of Grouping 385
R. F. JARRETT

Weight Given to Untenable Factors in Thinking 396
JOHN J. B. MORGAN

A Quantitative Study of Family Practices in Training Children in the Use of Money. 411
ESTHER ELIZABETH PREVEY

An Analytical Approach to the Study of Reading Skills 429
WILLIAM E. HALL AND FRANCIS P. ROBINSON

Equating High-School Intelligence Quotients with College Aptitude Test Scores 443
EDMUND G. WEBER

Publications Received 447

$6.00 per Year • Published Monthly September to May

WARWICK & YORK, INC.
BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF Educational Psychology

Established 1910

EDITED BY

JACK W. DUNLAP
University of Rochester

IN ASSOCIATION WITH

STEPHEN M. COREY
University of Chicago

JOHN G. DARLEY
University of Minnesota

BERTHA PETERSON HARPER
University of Rochester

HAROLD E. JONES
University of California

H. H. REMMERS
Purdue University

PERCIVAL M. SYMONDS
Teachers College (Columbia University)

PAUL A. WITTY
Northwestern University

H. E. BUCHHOLZ
Managing Editor

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. THE JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

Dr. Jack W. Dunlap is now on leave of absence from the University of Rochester and serving as Lieut. Commander in the U. S. Navy. For the 'duration', therefore, manuscripts, books and other materials for review, and correspondence regarding editorial matters should be addressed to H. E. Buchholz, Warwick & York, 10 E. Centre St., Baltimore, Md.

Manuscripts should be typed and double-spaced throughout including quotations, footnotes, and references. Return postage should be included with all unsolicited manuscripts.

THE JOURNAL is published monthly from September to May. The price per year in the United States and Pan-American countries is $6.00; $6.20 to Canada; and $6.40 to other foreign countries. Part-year subscriptions are 90 cents per issue ordered. Back volumes are $7.00 each, and back issues $1.10.

Subscribers should notify the Publishers of change of address at least four weeks in advance of publication of issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of next succeeding number.

WARWICK AND YORK *Publishers* BALTIMORE, MD.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXVI October, 1945 Number 7

ON THE PERMISSIBLE COARSENESS OF GROUPING

R. F. JARRETT

University of California

The psychologist has often to answer for himself and his students the very practical question, "How few class intervals may I employ with confidence that the errors of grouping in my mean and standard deviation shall not be excessive?" But when he seeks in textbooks on statistical methods for an answer to this query he is very likely to be disappointed and finds it necessary either to fall back upon some rule-of-thumb procedure for which he is unable to discover any rational justification, or, carrying his skepticism to the extreme, discard the grouped method of computing entirely and subject himself to the rigors of computation from ungrouped data.

It is perhaps too harsh to say that the answer to this question is not in the standard texts; it is just not made explicit. Fisher, for example, in his *Statistical Methods for Research Workers*, does indicate the answer, but in his usual dehydrated fashion. He does not present in its entirety the reasoning which underlies his suggested answer to the question, nor does he indicate how one might make use of the answer he provides in deciding before beginning computation whether or not the criterion he suggests is likely to be satisfied by the number of intervals it is decided to employ.

In the light of these deficiencies in textbook treatment of what may at times be a serious methodological problem to the psychologist, this note on the matter may be of interest to others.

It is well-known that in grouping data into intervals which are coarser than the units employed in measurement (or, indeed, in grouping into intervals imposed by the unit of measurement) a

certain amount of information is lost, and that this loss of information must be expected to result in inaccuracies in the statistics computed from such grouped data. It is generally understood, too, that these grouping errors are unsystematic in the case of the mean—*i.e.*, that they average to zero—but systematic in the case of the standard deviation, tending to make this statistic larger than it would have been were it computed from the ungrouped data. Sheppard's correction for this systematic error of grouping is at least mentioned in nearly every elementary text on statistical methods.

But it seems not so generally recognized that even though the effects of grouping errors upon the mean are 'unsystematic,' the standard deviation (or better, the variance) of the sampling distribution of means will be larger if those means are computed from grouped data than if they are computed from ungrouped data. This is clearly true, however, following at once from the independence of mean and grouping error in the samples. This phenomenon is precisely analagous to the increase in the variance of a distribution of 'true' scores when we add to each a random error of measurement such that the random errors average to zero and are uncorrelated with the true scores.

Specifically, were we to construct two sampling distributions of means by drawing a very large number of random samples of a given size from some specified population, computing the mean of each sample, first without grouping, and then after grouping into class intervals of size h, the sampling distribution of means computed from the grouped data would have the larger variance, although the means of these two sampling distributions would be identical. The means of the two corresponding sampling distributions of standard deviations would not, of course, be identical—whence Sheppard's correction—but in all other respects the case of the standard deviation is similar to that of the mean.

If each sample mean has added to it a random error of grouping which is uncorrelated with the error of random sampling, the variance of the sampling distribution obtained from grouped data is clearly the sum of two variances—the variance of random sampling and the variance of grouping. Thus the square root of the difference between the variance of the 'grouped' means and that of the 'ungrouped' means is the standard error of grouping (σ_G). That is, the standard error of grouping is the square root

of that part of the variance of the statistic which is due to errors of grouping.

Now in general the only reason for computing a statistic (*i.e.*, a numerical device descriptive of a sample) is that such statistics provide us with the only information available concerning the values of the corresponding parameters (*i.e.*, the corresponding numerical descriptions of the population from which the sample has been selected). The joint recognition of this fact and the fact of the increased variance of sample means computed from grouped data leads us at once to the criterion according to which we may decide whether or not our errors of grouping are excessive. Our estimates of the parameter are not completely dependable (we refer to one statistic as more 'reliable' than another) because of the errors of random sampling over which, for samples of a given size, we have no control whatever. And we have seen that grouping the data of our samples increases the variance of the statistic—makes it even less dependable as an estimate of the parameter. Thus our suspicion of coarse grouping is essentially concern over the possibility that the additional uncertainty as to the value of the parameter estimated from grouped data is too high a price to pay for the savings of effort in arriving at that estimate.

It is clear, then, that we may set a limit to the price we are willing to pay for computational time-saving. Fisher suggests[1] that the price is too high if it amounts to an increase of more than ten per cent in the standard error (the unreliability) of our statistic; or, in other words, the standard error of grouping (σ_G), he suggests, should not be greater than one tenth the standard error of random sampling (σ_S). This, of course, is not a dictum that one should never group so coarsely that the standard error of grouping is more than one tenth the standard error of random sampling, or that one should never require that the standard error of grouping should be less than one tenth that of random sampling. It is rather, like the 'level of confidence,' a matter for the choice of the individual in the light of his own temperament and the conditions surrounding the data at hand. We may at times be willing to permit σ_G to become as large as one fifth σ_S, and at other times we may be unwilling to permit it to

[1] *Statistical Methods for Research Workers*, 8th Edition, London: Oliver & Boyd, 1941, p. 49.

become as large as one twentieth σ_s. But clearly our decision as to whether or not grouping is too coarse must be made on the basis of this comparison.

If, now, we could evaluate the standard error of grouping, we would be in position to discover for each computational exercise whether or not our grouping is too coarse. If the ratio of σ_G to σ_s were greater than 0.1 (or whatever value had been decided upon), we should feel bound to recompute our statistics after grouping somewhat less coarsely! But this solution is not convenient as it stands. In order to determine the standard error of random sampling it is necessary for us to have an estimate of the standard deviation of the population, and this, in turn, requires that we have computed the standard deviation of the sample. That is, we must decide arbitrarily upon the number of intervals to be employed and after computation determine whether or not grouping error has excessively increased unreliability. What is needed is some device by which we may discover in advance how many class intervals to employ in order to be reasonably certain that our criterion will be satisfied when we apply the test at the conclusion of our computations.

Such a device is very simply constructed.

The standard error of grouping is first determined. The derivation of this quantity is beyond the scope of this note, but Fisher,[1] in a paper written for mathematicians, has shown that the standard error of grouping for both the mean and the standard deviation is approximately (ignoring terms of very small order):

$$\sigma_{G_{\bar{x}}} = \sigma_{G_s} = \frac{h}{\sqrt{12N}} \tag{1}$$

where σ_G is the standard error of grouping, $\bar{x}$ is the mean of the sample, s is the standard deviation of the sample, h the size of the class interval, and N the size of the sample.

The standard error of random sampling for the mean, it will be recalled, is:

$$\sigma_{s_{\bar{x}}} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \tag{2}$$

---

[1] "On the Mathematical Foundations of Theoretical Statistics." *Phil. Trans. Roy. Soc. London*, 1922, 222A, pp. 309–368.

where σ_s is the standard error of random sampling, σ is the standard deviation of the population sampled, and the other symbols have the meanings previously assigned them. A little algebraic rearranging shows that the ratio of the standard error of grouping (1) to the standard error of random sampling (2) in the case of the mean is:

$$c_x = \frac{h}{\sigma} \cdot \frac{1}{\sqrt{12}} \tag{3}$$

where we designate the ratio by c. That is, if we express h—the size of the class interval—in terms of the standard deviation of the population sampled, the ratio of the standard error of grouping to the standard error of random sampling is simply interval size times a constant.

We find c_s—that is, the ratio of the standard error of grouping to the standard error of random sampling for the standard deviation—in the same way. The standard error of grouping is here the same as for the mean, but the standard error of random sampling is somewhat smaller, being:

$$\sigma_{s_s} = \sigma_s = \frac{\sigma}{\sqrt{2N}} \tag{4}$$

and the ratio of (1) to (4) becomes:

$$c_s = \frac{h}{\sigma} \cdot \frac{1}{\sqrt{6}} = c_x \sqrt{2} \tag{5}$$

Now, if we could express the range of our sample in terms of the standard deviation of the population, we could easily express h in the same terms by simply dividing the range by the number of class intervals proposed. In general we can not express the range of our sample in terms of the population standard deviation, since we are in ignorance of that standard deviation; but thanks to the work of L. H. C. Tippett,[1] we do know the mean range (in these terms) of samples of all sizes from two to one thousand. Table XXII of Part II of Pearson's *Tables for Statisticians and Biometricians* gives the mean range of such samples expressed in terms of the standard deviation of the normal population from which the samples are randomly selected.

[1] "On the Extreme Individuals and the Range of Samples Taken from a Normal Population." *Biometrika*, 1925, 17, 364–387.

Thus, if we assume that our sample has a range equal to that of the mean range for samples of the same size, we may discover at once the ratio of the standard error of grouping to that of random sampling for any number of intervals which we might consider using.

Interval size would then be:

$$\frac{\bar{h}}{\sigma} = \frac{\bar{w}}{d} \tag{6}$$

where the bars above the letters designate means, w is the range of the sample expressed in units of σ, and d is the number of intervals to be employed.

Formula (3) thus becomes:

$$c_{\bar{x}} = \frac{\bar{w}}{d\sqrt{12}} \tag{7}$$

Thus before beginning our computations we may discover whether, on these assumptions (i.e., normality of supply, random sampling, sample range equal to mean range of samples of the same size) we may reasonably expect our criterion to be met. If we decide that σ_G should not exceed one tenth σ_x, we need only substitute in (7) the value of $\bar{w}$ (from Tippett's table) and the proposed value of d to discover whether or not the value of (7) is less than 0.1; in case of favorable outcome we may with confidence proceed to employ the proposed number of intervals. In the alternative case—if this value turns out to be greater than 0.1—we should know that the number of intervals proposed is not likely to be adequate.

However, a table which would show the value of this ratio for various sizes of samples and numbers of intervals would much facilitate the use of this criterion. A brief table of this kind is appended. As the writer envisages this table it will be used only for rough indications as to the value of the function tabled, so that an extended table is quite unjustified. The table here presented gives the value of $c_{\bar{x}}$, *i.e.*, the ratio of σ_G to $\sigma_{s_{\bar{x}}}$. If interest centers in keeping to a minimum the unreliability of the standard deviation, these values must all be multiplied by $\sqrt{2}$ as is apparent from (5).

Before considering the limitations of the table, we should call attention to some implications of the foregoing reasoning.

Fisher[1] observes that when making tests of significance it is important not to apply Sheppard's correction to the standard deviations on the basis of which the standard errors are estimated, but it is not obvious why the correction should not be used. It will be recalled that Sheppard's correction is necessary because on the average the standard deviations computed from grouped data are larger than those computed from ungrouped data. The correction involved is to subtract from the computed variance one twelfth the square of the size of the interval employed. Now, in tests of significance what we are doing is to enquire whether or not it is reasonable (on the basis of certain determinable probabilities) to believe that a discrepancy as large as an observed one may be accounted for solely in terms of the factors which we acknowledge as sources of undependability of the statistic. In the case of statistics computed from ungrouped data the only sources of unreliability are, in general, the variances of the populations involved and the sizes of the samples. But in the case of statistics from grouped data we must acknowledge the further source of unreliability—grouping error. Now, if we conduct our tests of significance as though we had no grouping error—*i.e.*, if we use as estimates of the population standard deviations the corrected estimates from grouped data—we shall in effect be refusing to acknowledge one of our sources of unreliability; we risk claiming significance where none exists.

We must, rather, obtain our estimate of the sampling error variance and add to it the grouping error variance; the square root of this quantity is the standard error required—the standard error obtained when we acknowledge grouping error to be a source of unreliability.

We remove the systematic effects of grouping upon sample variance by subtracting $\frac{h^2}{12}$ from the variance computed from grouped data:

$$s_c^2 = s_g^2 - \frac{h^2}{12} \tag{8}$$

where s_c is the corrected standard deviation, and s_g the standard deviation obtained from the grouped data.

The variance of the mean as deduced from this estimate of

[1] *Statistical Methods for Research Workers*, p. 113.

population variance is, then (ignoring for the sake of simplicity the bias of s—a matter unrelated to grouping error):

$$\sigma_{\bar{x}}^2 \doteq \frac{s_g^2 - \frac{h^2}{12}}{N} = \frac{s_g^2}{N} - \frac{h^2}{12N} \tag{9}$$

But the variance of the sampling distribution of means when they have been computed from grouped data is, as we have seen,

$$\sigma_{\bar{x}_g}^2 = \sigma_{\bar{x}}^2 + \frac{h^2}{12N} \tag{10}$$

and this is then approximately equal to

$$\left(\frac{s_g^2}{N} - \frac{h^2}{12N}\right) + \frac{h^2}{12N} = \frac{s_g^2}{N} \tag{11}$$

It is thus clear that we make both corrections (for systematic effect of grouping on s and for decreased reliability of the mean when obtained from grouped data) by employing the standard deviation obtained without applying Sheppard's correction, and the reason for Fisher's caution is apparent. It is clear, however, that when we wish to estimate the standard deviation of the population, Sheppard's correction should, of course, be used where its application is justified—where the variable is continuous and the frequency tapers to zero at both ends of the distribution.

A second implication of our reasoning is that if we hold constant the number of intervals, increasing N does not decrease the effect of grouping errors as might be expected from the known efficacy of increasing N in the case of most statistical ills. As N increases the range also increases, so that constant number of intervals is achieved with increasing N only at the expense of increasing their size and thus the standard error of grouping—and this at the same time that the standard error of sampling is decreased! This point is illustrated by any column of the table.

Before leaving the table we should consider the possible effect of failure to realize the assumed conditions. We have assumed that the population sampled is normal, that sampling is random, and that in every particular case the sample range is equal to the mean range for samples of the same size. Into how much error may we be led by depending upon the table when one (or more) of these assumptions is not met in our data?

We may dismiss at once consideration of the case in which sampling is not random. Sampling error theory makes no attempt to deal with non-random sampling. But we should consider the error of our table in the event the population sampled departs from normality. Very little is known of the sampling distribution of the range (upon which our reasoning hinges) for non-normal distributions, and that little concerns itself with very small samples—samples so small as to present no problem of grouping. What is known indicates that the standard error of the range—as well as the skewness and the kurtosis of its sampling distribution—is profoundly altered by changes in the form of the population sampled. The present state of our knowledge upon this point justifies only the caution that if the supply population be not fairly close to normal, the tabled values must be considered only as roughest approximations to the true values of the function.

We have assumed that the individual sample range is equal to the mean sample range for samples of that size, and in general this is not true. If the sample range is smaller than the mean range of samples of like size, the tabled values are conservative—the actual ratio of σ_G to σ_S will be smaller than the value appearing in the table. But the particular sample range may well be greater than the mean range, and we must enquire as to the values taken by (3) in the event the sample range were, say, at the 95th percentile in its sampling distribution—if it were so large that only five per cent of all ranges of samples of that size would exceed it. For it is clear that if a particular sample range is greater than the mean range of samples of its size, selection of interval size by reference to the table here presented will result in grouping which will be too coarse to satisfy the criterion. If we knew for each sample size the 90th, 95th, 99th percentiles of the sampling distribution of the range, we could prepare tables similar to the present one, showing the values of the ratio which would be expected on the assumptions that the sample range was so large as to be exceeded by only ten per cent, five per cent, or one per cent of sample ranges. But these percentiles are not known, so the value here tabled is the most reasonable one to present, since the mean range is accurately known and from this table we may arrive at approximations to the others.

MEAN RATIO OF STANDARD ERROR OF GROUPING TO STANDARD ERROR OF RANDOM SAMPLING (FOR MEANS) FOR SAMPLES OF VARIOUS SIZES AND FOR VARIOUS NUMBERS OF CLASS INTERVALS*

$$\frac{\sigma_G}{\sigma_{\bar{x}}} = \frac{\bar{w}}{d\sqrt{12}}$$

Sample Size	$\bar{w}$	Number of Class Intervals (d) 6	8	10	12	14	16	18	20	22	24	26	28	30
30	4.086	.197	.147	.118	.098	.084	.074	.066	.059	.054	.049	.045	.042	.039
40	4.322	.208	.156	.125	.104	.089	.078	.069	.062	.058	.052	.048	.045	.042
50	4.498	.216	.162	.130	.108	.093	.081	.072	.065	.059	.054	.050	.046	.043
60	4.639	.223	.167	.134	.112	.096	.084	.074	.067	.061	.056	.052	.048	.045
70	4.755	.229	.172	.137	.114	.098	.086	.076	.069	.062	.057	.053	.049	.046
80	4.854	.234	.175	.140	.117	.100	.088	.078	.070	.064	.058	.054	.050	.047
90	4.939	.237	.178	.142	.119	.102	.089	.079	.071	.065	.059	.055	.051	.047
100	5.015	.241	.181	.145	.121	.103	.090	.080	.072	.066	.060	.056	.052	.048
200	5.492	.264	.198	.158	.132	.113	.099	.088	.079	.072	.066	.061	.057	.053
500	6.073	.292	.219	.175	.146	.125	.109	.097	.088	.080	.073	.067	.063	.058
1000	6.483	.312	.234	.187	.156	.134	.117	.104	.094	.085	.078	.072	.067	.062

* When interest centers in the accuracy of the standard deviation, multiply each tabled entry by $\sqrt{2}$, or, approximately, increase each entry by 40 per cent.

Although the integrals of the sampling distributions of the range have not been computed, those of the correlation coefficient have been,[1] and these provide us with the means by which we may approximate the 95th percentiles of the range distributions. The procedure as outlined by Pearson[2] is to seek among the correlation sampling distributions for one having nearly the same β_1 and β_2 as those descriptive of the sampling distribution of the range and employ the integral of this distribution as an approximation to the integral required. Comparison of the Betas of the sampling distributions of ranges with those of the sampling distribution of correlation coefficients reveals that for $N = 100$ and in the vicinity of $\rho = 0.7$, 0.8, and 0.9, we find correlation sampling distributions with very nearly the Betas of the sampling distributions of the range for sample sizes which here concern us. Reference to David's excellent tables[3] reveals that for these distributions the 95th percentile is approximately 1.8 standard deviations removed from the mean of the distribution. For our purposes more accurate determination of this value is unnecessary. Knowing the standard deviation of the sampling distribution of the range in terms of the population standard deviation,[4] it is thus possible to determine by how much the numerator of (7)—and therefore the entire quantity—must be increased in order that the tabled value shall give the ratio of σ_G to σ_S under the assumption that the sample range lies approximately at the 95th percentile in its sampling distribution. Computation reveals that for this additional confidence in the table, the tabled values must be increased as follows: for samples of thirty, forty, and fifty, increase values about 30 per cent; for samples of sixty, seventy, and eighty, increase values about 25 per cent; for samples of ninety, one hundred, two hundred, and five hundred, increase values about 20 per cent, and for samples of one thousand, increase values about 15 per cent.

[1] David, F. N., *Tables of the Correlation Coefficient*, Biometrika Office, University College, London, 1938.

[2] *Tables for Statisticians and Biometricians*, Part II, pp. clvii ff.

[3] *Op. cit.*

[4] Tippett, *op. cit.*

WEIGHT GIVEN TO UNTENABLE FACTORS IN THINKING

JOHN J. B. MORGAN

Northwestern University

In a previous issue of this Journal* an experiment was reported which gave evidence that, when subjects are asked to assign relative ranks to equally untenable factors which have been lettered and presented in alphabetical array, their rankings will tend to correspond to this orderly array. This tendency on the part of our subjects was interpreted to mean that they were following the path of least resistance while going through the motions of reasoning.

The results of this first experiment have raised several significant questions. If an alphabetical sequence is not suggested in the experimental situation, will subjects more readily discern the fact that any ranking of untenable factors is illogical? What would happen were meaningless symbols used in place of the letters of the alphabet? Could it be that the subjects followed instructions and performed in an illogical manner merely to please their instructor?

In order to answer these questions, the essential part of the first experiment was repeated together with several modifications designed to determine what would happen when an easy arrangement was not suggested. In addition, the material was presented to high-school students by an outsider as well as by the instructor of a college class to his own students.

The complete test was as follows:

Which Button Rings the Bell?

Problem I. Suppose you have a series of nine buttons, each one designated by its own letter, as follows:

A B C D E F G H I

Suppose, further, that you know that one of these buttons, when pressed, will ring a bell. Your task is to discover which button

* "Following the path of least resistance in thinking," *J. educ. Psychol.*, 1944, 35, 27–38.

rings the bell by observing a series of situations in which several buttons are pressed at the same time.

Buttons D A H B are pressed simultaneously and the bell rings.
Buttons F C E A are pressed simultaneously and the bell rings.
Buttons B I A C are pressed simultaneously and the bell rings.
Buttons A D B G are pressed simultaneously and the bell rings.

You are to rank the different buttons in order, from 1 to 9, according to the importance you attach to each one as the probable ringer of the bell. Give a rank of 1 to the button you think most likely to have rung the bell by putting in the first space below (after 1) the letter of that button; after 2 put the letter of the button you think next likely to have rung the bell; after 3 put the letter of your third choice; after 4 put the letter of your fourth choice; and so on, until you have rated all nine of the buttons. When you have finished you should have a different letter in each space. The letter in space one represents the button you think most likely to have rung the bell, the letter in space nine represents the button you think least likely to have

TABLE I.—RESPONSES OF 157 STUDENTS TO PROBLEM ONE

Rank order	Buttons in Problem One								
	A	B	C	D	E	F	G	H	I
1	155	1				1			
2		154	3						
3		1	110	44	1				1
4	1		40	109	1	2	1	3	
5		1	1		105	17	7	23	3
6	1		2	1	14	113	10	8	8
7				1	20	12	106	9	9
8				1	9	8	8	111	20
9			1	1	7	4	25	3	116
Average ranks	1.05	2.02	3.32	3.81	5.68	6.09	7.19	7.35	8.49

TABLE II.—RESPONSES OF 157 STUDENTS TO PROBLEM TWO

Rank order	Buttons in Problem Two								
	D	H	E	G	C	I	B	A	F
1	154	2	1						
2	1	153		2				1	
3			96	60		1			
4			56	89	2	4	4	1	1
5			1	1	62	37	21	25	10
6		1		1	26	43	43	9	34
7	1	1	1		39	16	41	35	23
8	1		2	2	16	17	16	49	54
9				2	12	39	32	37	35
Average ranks	1.09	2.04	3.44	3.73	6.26	6.76	6.89	7.36	7.49

rung the bell. The letters in the intervening spaces represent a descending series of choices.

1____2____3____4____5____6____7____8____9____

* *

Problem II. This problem is to be done in the same manner as the one above. In this problem you have a series of buttons lettered as follows:

H I B D C E A F G

Buttons E G D H are pressed simultaneously and the bell rings.

Buttons C D G I are pressed simultaneously and the bell rings.

Buttons D H E B are pressed simultaneously and the bell rings.

Buttons F A D H are pressed simultaneously and the bell rings.

Do just as you did in the problem above, that is, rank the buttons in order from the one most likely to have rung the bell to the one least likely to have rung it.

1____2____3____4____5____6____7____8____9____

* *

Problem III. This, and the following problems, are to be done in the same manner. The only difference is that each problem has the nine buttons labelled in a different manner, sometimes with letters and sometimes with other symbols.
Suppose you have a series of buttons labelled as follows:

U M X K O S Y R T

Buttons K T S M are pressed simultaneously and the bell rings.

Buttons S R K Y are pressed simultaneously and the bell rings.

Buttons U M X S are pressed simultaneously and the bell rings.

Buttons K S R O are pressed simultaneously and the bell rings.

Rank the buttons in order from the one most likely to have rung the bell to the one least likely to have rung it.

1____2____3____4____5____6____7____8____9____

* *

TABLE III.—RESPONSES OF 157 STUDENTS TO PROBLEM THREE

Rank order	Buttons in Problem Three								
	S	K	M	R	X	U	T	Y	O
1	156	1							
2	1	154	2						
3		1	124	32					
4		1	31	118		2	1	2	2
5				3	20	48	62	14	10
6				1	45	22	16	50	23
7					22	48	19	19	49
8				1	45	21	11	52	27
9				2	25	16	48	20	46
Average ranks	1.01	2.01	3.18	3.92	6.44	6.54	6.77	7.05	7.44

Problem IV. Suppose you have a series of buttons labelled as follows:

& @ * ? ¢ ½ % # $

Buttons * ¢ ½ ? are pressed simultaneously and the bell rings.

Buttons # @ * % are pressed simultaneously and the bell rings.

Buttons $ * # ¢ are pressed simultaneously and the bell rings.

Buttons % # & * are pressed simultaneously and the bell rings.

Rank the buttons in order from the one most likely to have rung the bell to the one least likely to have rung it.

1____2____3____4____5____6____7____8____9____

* *

TABLE IV.—RESPONSES OF 157 STUDENTS TO PROBLEM FOUR

Rank order	Buttons in Problem Four								
	*	#	¢	%	½	@	?	&	$
1	156	1							
2		153	2	1			1		
3	1		96	56		1		3	
4		2	54	93	2	1			5
5			1		58	18	16	45	19
6			2	2	15	47	55	17	19
7		1	2	2	16	57	41	15	23
8				1	53	22	12	23	46
9				2	13	11	32	54	45
Average ranks	1.01	2.05	3.43	3.78	6.63	6.70	6.89	7.07	7.41

Problem V. Suppose you have a series of buttons labelled as follows:

$? * ¢ % & # @ ½

Buttons ¢ ? $ * are pressed simultaneously and the bell rings.

Buttons * # ? $ are pressed simultaneously and the bell rings.

Buttons $ ¢ & @ are pressed simultaneously and the bell rings.

Buttons ? $ % ½ are pressed simultaneously and the bell rings.

Rank the buttons in order from the one most likely to have rung the bell to the one least likely to have rung it.

1____2____3____4____5____6____7____8____9____

* *

TABLE V.—RESPONSES OF 157 STUDENTS TO PROBLEM FIVE

Rank order	Buttons in Problem Five								
	$	?	*	¢	&	%	#	@	½
1	155		1						1
2		152	1	4					
3		3	98	55			1		
4		1	54	91	2	2	3	3	1
5		1	1	2	23	60	56	10	4
6			2		103	22	7	16	7
7	1			2	11	24	57	47	15
8				1	10	43	12	56	25
9	1			2	8	6	21	15	104
Average ranks	1.09	2.05	3.38	3.74	6.17	6.40	6.51	6.82	8.34

Problem VI. Suppose you have a series of buttons labelled as follows:

V U T S R Q P O N

Buttons T V O U are pressed simultaneously and the bell rings.

Buttons V P S N are pressed simultaneously and the bell rings.

Buttons U Q R V are pressed simultaneously and the bell rings.

Buttons S U V T are pressed simultaneously and the bell rings.

Rank the buttons in order from the one most likely to have rung the bell to the one least likely to have rung it.

1____2____3____4____5____6____7____8____9____

* *

TABLE VI.—RESPONSES OF 157 STUDENTS TO PROBLEM SIX

Rank order	Buttons in Problem Six								
	V	U	T	S	P	Q	R	O	N
1	154	3							
2	2	151	3	1					
3	1	2	108	43	1	1			1
4			44	105			2	5	1
5			1	3	23	12	65	44	9
6				2	42	78	16	6	13
7				1	71	21	13	15	36
8			1	1	17	39	24	64	11
9		1		1	3	6	30	23	86
Average ranks	1.03	2.04	3.31	3.83	6.56	6.65	6.66	7.00	7.92

Problem I corresponds exactly with a problem used in the experiment reported earlier. In this problem both the position of letters on the keys and the frequency of appearance in the bell ringings suggest an orderly, alphabetical sequence. Button A appears in all four ringings, button B appears in three, buttons C and D each appear in two, and the remaining five buttons —E F G H I—each appear but once.

Problem II uses the same letters as does Problem I, but neither the arrangement on the keys nor the frequency of appearance corresponds with any common alphabetical order. Button D appears in all four ringings, button H appears in three, buttons E and G each appear in two, and the remaining five buttons—I, B, C, A, F—scattered about on the keys, appear but once each.

Problem III utilizes letters not usually found in enumerative sequences and with no definite relation between the order on the keys and frequency of appearance. Button S appears in all four ringings, button K appears in three, buttons M and R appear in

two each, and the remaining five buttons—U, X, O, Y, T—appear in but one ringing each.

Problem IV was designed to represent the least probability of any suggestion of meaning or order. The symbols were chosen at random and the order on the keys does not correspond with frequency of appearance. Button * appears in all four ringings, button # appears in three ringings, buttons ¢ and % appear in two each and the remaining buttons—&, @, ?, ½, $—appear but once each.

Problem V uses the same symbols as those used in Problem IV but with the arrangement on the keys corresponding with the frequency of appearance. Button $, on the left end of the bank of keys, appears in all four ringings; button?, next in order, appears in three ringings; buttons * and ¢, occupying positions three and four, appear in two ringings each; and buttons %, &, #, @, and ½ appear but once each.

Problem VI uses nine letters from the latter part of the alphabet in a reverse sequence both as to order on the keys and frequency of appearance. Button V appears in all four ringings, button U appears in three, buttons T and S appear in two each, and the remaining five—R, Q, P, O, N—appear but once each.

The test was given to eighty-five students in a class in psychology at Northwestern University and to seventy-two students at the New Trier High School in Winnetka, Illinois.*

RESULTS

The results for the entire group of one hundred fifty-seven students in all six problems are given in Tables I to VI, one table for each problem.

In each table the columns are arranged so that the average ranks increase from left to right. It will be observed that the button which is given the highest rank in each problem is the one which appeared in all four ringings of the bell; that the button given the second position is the one appearing in three ringings, and that the buttons given third and fourth places were those appearing twice.

In the previous experiment two subjects (one per cent) stated

* The writer is grateful to Miss Eileen Duberstein who gave the tests at the High School, to Mr. Matthew P. Gaffney, Superintendent of the New Trier High School, and to his staff, for their help.

that it was not logical or possible to rank the remaining buttons after it had been demonstrated that the bell rang when they were not pressed. In the present experiment none of the students objected. Thus the present study confirms one of the conclusions stated in the previous report, namely; there is a tendency to weight a factor according to the frequency of its appearance in spite of the fact that it has been shown to have no value. It should be noted that our respondents were told in advance that one button would ring the bell; then it was demonstrated that the bell would ring in the absence of all but one button. In spite of such evidence, the students ranked all the buttons, giving more weight to those which appeared more often.

TABLE VII.—RANKINGS OF BUTTONS IN PREVIOUSLY REPORTED STUDY COMPARED WITH RANKINGS FOR IDENTICAL PROBLEM IN THIS STUDY

	Buttons								
	A	B	C	D	E	F	G	H	I
Av. rankings first study	1.00	2.01	3.13	3.84	5.40	6.25	7.07	7.56	8.68
Av. rankings this study	1.05	2.02	3.32	3.81	5.68	6.09	7.19	7.35	8.49

Analysis of Variance

Source of variation	Degrees of freedom	Sum of squares	Mean square
Total	17	108.7564	
Between present and previous study	1	0.0007	0.0007
Between buttons	8	108.6373	13.5797
Interaction	8	0.1184	0.9148

Ratio (F) between studies/interaction equals 0.047, nonsignificant.
Ratio (F) between buttons/interaction equals 917.5, highly significant.

Since Problem I was an exact duplicate of one previously reported, a comparison of results from the two was made as a check upon reliability. This comparison is displayed in Table VII. The rankings of the buttons in the two experiments are remarkably similar. An analysis of variance of the two sets of rankings reveals that the ratio of the variance attributable to the difference between the two groups of subjects to the variance attributable to experimental error (F ratio) is 0.047. Practically all the variance is due to differences between the buttons.

TABLE VIII.—COMPARISON OF 72 HIGH-SCHOOL WITH 85 COLLEGE STUDENTS*

	Problems compared					
	1	2	3	4	5	6
Variance due to difference between groups....	0.00013	0.0043	0.0072	0.0122	0.0697	0.0021
Variance due to experimental error.	0.01700	0.0123	0.0181	0.0839	0.4698	1.3048
F ratios.....	0.00765	0.3500	0.3980	0.1453	0.1484	0.0016

* The F ratios sustain the hypothesis that there was no significant difference between the two groups in their responses to this experiment. With eight degrees of freedom, it would require an F ratio of 5.32 or more before the null hypothesis could be rejected.

Another check on reliability was made by comparing the rankings of the high-school students with those of the college students in all six of the problems in the present study. The results of these comparisons are given in Table VIII. There was no significant difference between the two groups in any of the problems. These findings provide evidence for a negative answer to one of the questions raised: Did the subjects in the previous experiment perform in an illogical manner merely to please their instructor? The high-school students had never

seen the person who gave the tests. The only pressures applied were those inherent in the structure and directions in the test.

What determined the ratings given to those buttons which appeared but once in the bell ringings? A survey of the papers suggested four different criteria which might have been used, namely; (a) the order in which the letters or symbols appeared on the keys, (b) an alphabetical order, (c) the order in which they appeared when read off by horizontal rows, and (d) the order in which they appeared when read by vertical columns.

In order to determine the part played by each of these factors, a product-moment correlation was computed between the average rank order which would have resulted had one of these criteria been followed implicitly and the actual rank order which was obtained from the responses of our subjects. The method may be illustrated by considering Problem Three. The five buttons which appeared but once were X, U, T, Y, O. The actual average rankings of these five buttons were: X – 6.44; U – 6.54; T – 6.77; Y – 7.05; and O – 7.44. Had our subjects been controlled purely by the order on the keys the average rankings would have been: U – 5; X – 6; O – 7; Y – 8; and T – 9. The correlation between these two is .13 ± 30. Similarly, to test the influence of alphabetical sequence, the hypothetical ratings of O – 5; T – 6; U – 7; X – 8; and Y – 9 were correlated with the actual ratings of O – 7.44; T – 6.77; U – 6.54; X – 6.44; and Y – 7.05. This correlation was –.43 ± 25. Similar procedures were used to test the influence of the order of appearance of the letters by columns and rows in the presentation arrays. The correlations thus found are brought together in Table IX. Since only six items are involved the probable errors of these correlations are quite high.

When the order on the keys followed an alphabetical sequence, as in Problem I, and when A was the key appearing in all four ringings, B the key appearing in three ringings, C and D the keys appearing in two ringings, and E, F, G, H, and I appearing in but one ringing each, it is clear that our subjects followed faithfully this arrangement in their ratings. The correlation of .98 demonstrates the closeness of the approximation of actual ratings with the hypothetical rating of E – 5; F – 6; G – 7, H – 8, and I – 9. There was no significant tendency in

Problem I for our subjects to follow the order of presentation arrays either by rows or columns.

Thus, it seems clear that subjects can be led to arrange non-logical material in a definite sequence if some definite order is suggested by the arrangement of materials. The application of this principle to ordinary situations is apparent. If one has

TABLE IX.—PRODUCT-MOMENT CORRELATIONS BETWEEN ACTUAL AVERAGE RATINGS AND THE AVERAGE RATINGS WHICH WOULD HAVE BEEN MADE HAD CERTAIN RATING METHODS BEEN FOLLOWED IN RANKING THE FIVE BUTTONS EACH OF WHICH APPEARED BUT ONCE

Method of ranking	Problem					
	1	2	3	4	5	6
In order of position on keys	.98 ± .01	.62 ± .19	.13 ± .30	.31 ± .27	.82 ± .10	.86 ± .08
In alphabetical order	.98 ± .01	−.19 ± .29	−.43 ± .25			−.13 ± .30
In order of appearance in rows	.30 ± .27	.94 ± .04	.28 ± .28	.70 ± .15	.71 ± .15	−.16 ± .29
In order of appearance in columns	−.09 ± .30	.17 ± .29	.81 ± .10	.17 ± .29	.79 ± .11	.76 ± .13

the desire to make it appear that certain responses are determined by logical processes, when there is no logical reason for preferring one arrangement rather than another, one can be practically certain that one will get the desired response if one presents the materials in such a manner that the desired arrangement is clearly suggested.

What happens when there is no obvious suggestion of an alphabetical arrangement? Problem II was designed to give an answer to this question. The same letters were used as in Problem I, but they were arranged in a random manner. The

guiding principle which seemed to hold with our subjects in this problem is the order of appearance in rows, namely; the order C I B F A, which is strikingly close to the order obtained from our subjects, that is, C I B A F.

Problem III utilized letters which are not customarily used in enumerative discourse and these letters are arranged in random order. The only significant correlation is with the order of appearance of these keys as read by columns in the presentation array. The actual rankings by our subjects was X U T Y O; whereas if arranged in order of appearance by columns the order would be U T X Y O. The product-moment correlation between the column order and the obtained order is .81 ± .10. Why our subjects favored an arrangement approximating the appearance by rows in Problem II and an arrangement approximating the appearance by columns in Problem III we do not know.

Problem VI is the only other problem using letters. In this problem nine letters from the latter half of the alphabet were used in reverse alphabetical order. From the correlations shown in Table IX it is clear that the subjects were more influenced by the position on the Keys than by the alphabetical sequence. This suggests that spatial order is more important than alphabetical sequence when the letters are from the latter part of the alphabet.

Problem V has an arrangement of symbols which follows the same pattern as Problems I and VI. This problem was designed to determine how closely our subjects would approximate the positional sequence without any help or hindrance from alphabetical arrangements. The highest correlation, .86 ± .08, in this problem is with spatial arrangement of the keys. Thus, we have support for the hypothesis that spatial arrangement plays a large part in the responses of our subjects in this experiment.

Problem IV was designed to contain a maximum of disorder and a minimum of any meaning which might lead to definite rankings. While there is some tendency to follow appearance in rows, it is not too striking and all the other correlations are low.

An examination of the way in which various subjects performed the task involved in this experiment throws additional light on the results obtained. From markings on some of the papers and from questioning of a representative group of subjects after the work was done, it was found that there were two general methods which predominated. The first was to read through the sixteen

letters (or symbols) displayed in the different ringings of the bell and to make a tally mark by the button carrying the label with each appearance. The result of such a procedure in Problem I, for example, would be four tallies at button A, three tallies at B, two tallies at each of C and D, and one tally at each of the remaining buttons. The easiest way to record the buttons in the answer blanks would then be to copy them in an order corresponding to the position on the keys. The high correlations between the position on the keys and the actual rankings in Problems I, V, and VI suggest that many subjects used this method.

Other subjects would scan the array of letters (or symbols) in the bell ringing displays in order to select the one appearing in all four ringings and mark it; then search for the button appearing three times and mark it, and so on. Having marked the array, some would read off the results by columns and some by rows.

A reasonable interpretation of these findings is that our subjects were more influenced by the mechanics of doing the particular tasks than they were by the logical procedures underlying the work. Spatial arrangements which made it easier for a particular procedure resulted in a more uniform usage of that method. Another incidental feature supports this conclusion. In Problem V the symbol ½ occupies the last position on the bank of keys and in the presentation array; it received an average rating of 8.34, reflecting the combined influence of position on the keys and both columnar and row reading.

SUMMARY AND CONCLUSIONS

With the advance information that one button in nine would ring a bell, subjects were given the problem of selecting the correct button. They were not permitted to note the effect of pushing each button separately; instead they knew only that the bell rang when certain combinations of buttons were pressed. In each of the six problems in the experiment a certain button appeared in all four ringings, another button appeared in three, two buttons appeared twice, and each of the remaining five appeared but once in the ringings. Six variations of the problem were given, one of which was the exact duplicate of an arrangement used in a previously reported experiment.

1) When students are given an inductive problem in which the reasoning universe is definitely limited, they tend to proceed in a manner that would be logical only in an unlimited reasoning universe. When it is known that one of a number of factors is the potent one and the task is to discover the identity of that one factor, one negative instance is enough to exclude any factor from consideration. When the given consequence occurs without the presence of a given factor, that one occurrence without its presence is enough to warrant its exclusion from consideration. In spite of this simple fact, our subjects ranked ineffective factors in order of the frequency of their appearance. This finding confirms the conclusion of a previously reported study.

2) Students will also assign differential ratings to useless factors each of which appear but once in association with the potent one. These ratings can be influenced by the way in which the material is presented. A clear-cut spatial arrangement, reinforced by alphabetical sequence, provides the most influential means for controlling such ratings; but spatial order remains very effective when the help of alphabetical sequence is removed. When no suggestions are contained in the way the materials are presented, subjects show greater diversity in their responses but they continue to give differential ratings.

3) These results suggest a principle which is important both in persuasion and in guarding ourselves against irrational pressures. A factor does not need to be logically distorted in order that it may receive apparent logical support or apparent logical refutation. All that seems necessary is to put the irrational factor in such a setting that the person to be influenced is given the impetus to proceed in a certain direction or manner and the momentum thus engendered will carry him through to a conclusion about which he ceases to be intelligently critical. Instruct a person to use logic, but get him involved in some mechanical complication, and he soon forgets the logic in working on the details of the problem. A more general hypothesis growing from this experiment might be stated as follows: It may be that crooked thinking comes not so much from the fact that the pathway to truth is filled with obstacles as from the absence of barriers on the highway to error.

A QUANTITATIVE STUDY OF FAMILY PRACTICES IN TRAINING CHILDREN IN THE USE OF MONEY*

ESTHER ELIZABETH PREVEY

Director, Family Life Education, Kansas City, Missouri, Public Schools

REVIEW OF LITERATURE

In surveying the literature on the question of children and money, it was found that two general types of material are available. On the one hand, a number of authorities interested in child growth have written extensively on the problem of how children should be trained in the use of money. On the other hand, there are reports of the few research studies which have been made in this area.

Gruenberg and Gruenberg,[14] Fisher,[10] Andrews,[44] Eliot,[9] Kirkpatrick,[22] Ojemann,[30] and others have stressed the educational values in supplying children with money which they must manage. They have pointed out that adult financial problems may be encountered in miniature with the spending of the childhood allowance. Also, the values in making choices and decisions and the acceptance of responsibility are emphasized.

Research studies on children and the use of money are not numerous. Kohler,[23] Dismoor,[8] the White House Conference Studies,[43,44,45] and Morgan[29] have gathered data on children's allowances and earnings. The actual percentage of children receiving allowances varies from one study to another. One conclusion seems to be warranted; namely, that there appears to be an increasing tendency for parents to supply children with money in some systematic manner rather than 'doling' it out as needed. In the matter of earnings, there is little specific information. Again, the percentage of children earning varies with each study. There seems to be, however, a decreasing tendency for children to have their own money by earning it.†

* Abstract of a Thesis submitted to the Graduate Faculty of The University of Minnesota in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, July, 1944. The original thesis is filed in the library of the University of Minnesota, Minneapolis, Minn.

† As all of the studies which have been done were completed before the present war, this result is probably not in accordance with conditions as they are today.

Hanson[18] concluded from her study that receiving money according to a well-defined plan with parental guidance was more favorable to child development than haphazard methods. Her group was small and there was a wide lage range and socio-economic spread. Her study appears to be more of an exploratory type than one in which reliable results have been obtained.

A number of studies[22,31,39,41,42] have been made on family budgets. None of these give information on the problem of training children in the use of money.

There is practically no information regarding the spending responsibilities which children assume.

There are large areas about which there are few or no data. There are few facts about children's experiences in saving, borrowing, lending, giving, or in sharing information about the family's financial resources and obligations. We know little about the relationship between parental practices in teaching children the use of money and the later money habits developed by the children as they near maturity. Also, there is little information about the relationship between adjustment and utilization of one's economic resources.

Most of the investigations have been on the basis of securing data from the children with little emphasis on what the parents themselves are doing. The younger age range has been studied more than older groups of children. There are few or no studies on either a limited age range or a restricted economic group. With these considerations in mind the present investigation was planned.

THE PROBLEM

This investigation had two chief objectives. First, it was aimed to determine on a quantitative basis information about (a) the methods which parents from the upper socio-economic levels are following in teaching children the use of money and (b) the relationship between childhood experiences with money and later money habits. Second, it was aimed to determine the relationship between the responsibility which children are allowed to assume in the use of money and their general level of adjustment and the degree of self-reliance which they attain during their high-school years.

In seeking to determine the methods which parents use in teaching children the use of money, the following areas were con-

sidered: (a) methods of supplying children with money, (b) amounts of money children have to manage, and (c) children's experiences in spending, giving, saving, borrowing, lending, budgeting, earning, and problems of family finance. These areas were considered from the standpoint of the principles or practices which the parents are using with the children.

The relationship between the responsibility which children are given in the use of money and their general level of adjustment and state of self-reliance or independence was investigated on the theory that there might be more or less relationship between parent practices with children in the use of money and the children's home, social and emotional adjustment, as well as their level of independence in other areas of behavior.

The method of study.—A homogeneous group of parents and children from the upper socio-economic level was used for the following reasons, (a) the income would be sufficiently large to supply the children with money, (b) the parents would be above the average intellectually and likely would have educational advantages which might make it possible for them to have developed a well-defined philosophy regarding children's training in the use of money, and (c) no study has been done on this problem which has centered its attention on the upper socio-economic level.

The types of data secured for each subject were: (a) information about parent practices employed in teaching children the use of money, (b) general information regarding age, educational status and birthplace of the parents, family income, father's occupation, the size of family, (c) Bell Inventory scores on each child, (d) self-reliance scores on each subject secured through the use of Stott's Self-Reliance Scale. In addition to the above mentioned information, it was possible to secure data regarding the later money habits developed by fifty of the subjects through a follow-up study made four years after the original data were secured.

The Bell Inventory[4] was selected as a measure of adjustment. Through the use of this scale scores can be secured on adjustment toward the home, attitudes toward health factors, social aggressiveness and emotional stability, as well as a total score which might be considered as a measure of general adjustment.

Stott's Self-Reliance Scale[37] was used for measuring self-reliance. It is divided into subscales which give scores on several

types of self-reliance. The subscales are referred to by Stott as 'Factors.' Factor I has to do with the tendency to want to make one's own decision and to rely upon one's own judgments particularly in regard to matters of a personal nature. It might be called Independence of Decision in Personal Matters. Stott labels it simply 'Independence.' Factor II* is not used in this study. Factor III is concerned with resourcefulness, dependability and willingness to work in a group situation and is called 'Social Resourcefulness' or Resourcefulness in the Group Situation. Factor IV is concerned with the dependable and responsible type of person who is conscientious in keeping his agreements, meeting his obligations and "keeping up his end" in relationships with others. It is called 'Personal Responsibility.'

Each child rated himself on the Bell Inventory and the Stott Self-Reliance Scale. In addition, each mother rated her own child on the Stott Scale.

The data on parent practices in teaching children the use of money were secured from the mothers by means of an interview using a specific form.† Information was secured about: (a) the methods which the parents were using to supply their children with money, (b) the experiences which they were offering to their children in spending, giving, lending, borrowing, saving, and earning, (c) the responsibility which they were giving to the children in planning the use of money and the type of guidance offered, (d) the opportunities which were provided for helping children to understand household expenses and the family's financial status and problems encountered and their solutions. The interview was conducted on the basis of free conversation with the mother, the investigator asking questions when necessary to secure the desired information. The information secured will be referred to as 'Parent Practices with Children.'

The enthusiastic response and interest shown by the mothers during the interview was most gratifying. On the whole, information was given freely and willingly, accompanied by an

---

* Factor II is not used because it appears to bear little relationship to this study.

† The interview form was devised by the investigator on the basis of a survey of the literature plus practical experience in Family Life Education. It was then submitted to several authorities in child development and family life and revised according to their criticisms. The second form was used in a preliminary study of six cases, following which it was revised a second time.

attitude of sincerity which indicated a desire to give accurate data.

The follow-up study which was made four years after the original data were collected had as its object the determination of the money habits which the subjects actually developed as they neared maturity. Data were secured on fifty of the original subjects by means of a questionnaire mailed to the mothers. It was hoped that, through the follow-up study, it could be determined what relationship existed between the practices used by the parents in teaching children the use of money and the actual habits which these individuals later developed. The data secured through the follow-up study will be referred to as 'Later Money Habits of the Subjects.' A more detailed description of the follow-up study is given later.

Selection of families for study.—From the entire Senior class* (one hundred ninety-seven boys, one hundred ninety-four girls) in a high school situated in the upper socio-economic residential district of a midwestern city of approximately half a million population, a homogenous group was selected for study on the following basis: (a) parents were living together and the student was living with them, (b) both parents had attended high school or college, (c) the fathers were of the Professional or Managerial Group when classified according to the Minnesota Occupational Scale,[12] (p. 501), and (d) the students were within the age range fifteen years, six months to seventeen years, six months.

On this basis, one hundred forty-three families qualified as possible cases for investigation. Of these, it was found that one hundred could actually be included in the study. Forty-three cases were not used because the mothers did not care to participate, or because, for one reason or another, participation was impossible. As very few cases were rejected because of lack of interest or antagonism it seemed highly probable that the group studied was a good sample.

Description of the families studied.—A group of fifty boys and fifty girls, fifteen one-half to seventeen one-half years of age, living with both of their parents and in their senior year of high school were used as the subjects. Their fathers were from the upper socio-economic level, all belonging to groups I or II accord-

* In this city there is an eleven-year public school system, seven years in the elementary school and four in the high school. The children are, therefore, on the whole one year younger than the majority of high-school seniors in other cities.

ing to the Minnesota Occupational Scale. All of the parents had had at least one year of high-school education while fifty-eight per cent of the fathers and forty-seven per cent of the mothers had had some college education. The median family income was $4843.75. Ninety-nine per cent of the mothers and ninety-six per cent of the fathers were American-born. Sixty-four per cent of the families were two or three children families.

The follow-up study on later money habits of subjects.—Four years after the original data were collected a follow-up study was made, the aim being to determine, if possible, the relationship between parent practices in teaching children the use of money and the money habits which the subjects actually developed as they neared maturity.

The follow-up study was made by means of a questionnaire mailed to the eighty-four families who could be located. Fifty-nine replies were secured of which fifty (twenty-seven boys, twenty-three girls) were complete enough to use. The data secured through this follow-up study will be referred to as the 'Later Money Habits of the Subjects.' All of these young people (ages nineteen one-half to twenty-one one-half years at the time of the follow-up study) had incomes of their own to manage.

QUANTITATIVE TREATMENT OF THE DATA

In order to analyze the data on parent practices in teaching children the use of money, qualitative facts secured through the interviews with the mothers were translated into quantitative terms. This was done by selecting twenty items descriptive of practices which parents were using with children and submitting them to a group of qualified judges* for rating in order of importance in training children in the use of money. Six items were discarded because of marked disagreement among the judges on their relative importance. Each of the remaining items was given the median value of the judges' ratings. Under 'Parent Practices' is given the list of these fourteen items and their values. Each case was scored on the presence or absence of each of the fourteen practices described. The total score of all items for a particular case was considered as the score for that subject.

* Returns were received from thirty-three judges all of whom are widely recognized as authorities in the field of Child Development and Family Life Education.

PARENT PRACTICES*

	Value
1) The child is provided with money according to a well-defined and understood plan.	49
2) The child's own money is a small amount expected to cover minor expenses only—'spending money.'	13
3) The child's own money is an amount expected to cover minor expenses plus a few other items of expense such as recreation, club dues, or minor articles of clothing.**	26
4) The child's own money is an amount expected to cover practically all of his usual expenses including most of his clothes.	40
6) The child (as a rule) must accept the consequences of unwise planning or spending rather than receiving additional money when he is 'short.'	48
7) The amount of money received by the child is adequate to cover the items of expense which it is expected to cover.	45
8) The child received guidance in the use of his money—guidance being interpreted to mean that the parents may help him to plan but actual decisions are left to the child.	50
11) As the child's income has been increased, his responsibilities in planning its use and in spending have also been increased.	48
13) The child is encouraged to have earning experiences.	48
14) The child is encouraged to save for definite goals to be achieved in the near future.	46
15) The child is encouraged to save just for the sake of saving —indefinite goals to be achieved at an indefinite time.	11
16) The child is encouraged to save for definite goals to be achieved in the distant future. For example, life insurance or travel after the war.	23
17) The child is given opportunities to borrow and loan on a business-like basis. Such opportunities are of a type which would lead to an understanding of the principles of investment.	42
20) The child has opportunities to share in the family financial planning and to develop an understanding of the family financial problems.	49

Data on the subjects' later money habits as determined by the follow-up study.—The data secured through the follow-up study were also assigned quantitative values determined in the following manner:

---

* Items 5, 9, 10, 12, 18, and 19 are omitted because they were the discarded items.

** Each subject was given credit for one of 2, 3, or 4, but not more than one.

The questionnaire had been compiled around broad areas in the individual's use of money, for example, saving habits, budgeting the use of money, thrift, and several others. Under each of these there were three to five statements describing possible habits which the individual could have developed. The percentages of positive answers for each statement were computed and values were assigned in terms of deviations from the mean in a normal curve using the procedure as described by Kelley[21] (p. 100). It was found that eight items from the follow-up questionnaire could be used for evaluating the later money habits of the subjects. These were:

LATER MONEY HABITS OF SUBJECTS

I. Saving Habits

	Value
1) He usually has some fairly important purpose for which he saves over long periods of time.	4.202
2) He usually saves for some definite purpose but for periods of time of less than one year.	3.129
3) He usually saves some money but has no definite goal.	2.408
4) He plans to save for some specific purpose but is likely to spend his savings for something else before the goal is reached.	1.863
5) He usually spends his money as it is received. Seldom saves.	1.142

II. Contributions A

1) He feels a responsibility for making contributions to Red Cross, War Chest Funds, etc., therefore, contributes as much as he is able.	3.218
2) He contributes small amounts to those funds in order to say he has made a contribution.	1.582
3) He seldom contributes to these funds.	.903

III. Contributions B

1) He contributes regularly to church.	3.757
2) He contributes occasionally to church.	2.461
3) He seldom or never contributes to church.	1.380

IV. Personal Responsibility for Finances

1) He determines for himself how his money shall be spent.	3.966
2) He determines to a large degree how his money shall be spent, receives some advice from parents.	2.417
3) He seldom makes expenditures without consulting parents.	.579

V. Budgeting the Use of Money

	Value
1) He plans the use of his money carefully in accordance with his needs and income.	3.416
2) At times he plans carefully, at times he is careless in planning the use of his money.	1.947
3) He seldom bothers to plan the use of his money—just spends.	1.022

VI. Financial Control

1) He not only plans the use of his money, but he usually makes his money cover the necessary expenses over periods of time.	3.497
2) He plans well but is not able to live up to his plans. He tends to spend as ideas or needs arise regardless of previous planning.	2.355
3) He spends as needs arise but seldom has to borrow.	1.654
4) He is frequently in debt because he has to borrow from one pay day to the next to keep going.	0.000

VII. Thrift

1) He is reasonably thrifty in the use of money—without being careless or miserly.	3.380
2) He is apt to be careless in spending, not evaluating his needs.	1.894
3) He is exceedingly careful in his spending, not spending five cents if he can avoid it.	1.015

VIII. Expenses Which Come Out of His Own Money

1) His own money covers all of his expenses.	4.218
2) His own money covers all of his minor expenses plus some major item of expense such as clothes.	2.938
3) His own money covers all of his minor expenses but no major expenses such as clothes or board and room.	1.672

The total score for each subject was secured by summing the values received for the above described items, calculating the average score for each subject, correcting it to two decimal places and removing the decimal for convenience in further calculations.

Data from the Bell Inventory and Stott Self-Reliance Scale.—The Bell Inventory was scored according to Bell's directions using the unweighted scores. This gave five scores for each subject. These scores represented (a) adjustment toward the home, (b) adjustment toward health, (c) social adjustment, (d)

emotional adjustment, and (e) total adjustment. Low scores on this inventory are indicative of 'good' adjustment.

The Scott Scale was also scored according to the directions given by Stott using weighted values. This gave two scores for each subject in each of the following areas; (a) independence (Factor I), (b) social resourcefulness (Factor III), and (c) personal responsibility (Factor IV). One set of three scores was determined by the subject rating himself and the other set by the mother rating the child. In this scale, high scores are 'good.'

RESULTS

Parent practices in teaching children the use of money.—The practices used by parents in teaching children the use of money were scored as described above. The highest possible score on this basis was 499. Individual scores ranged from 0 to 450. The mean score for the boys was 238 (SD 107.5), for the girls 175, (SD 112.5). It will be noticed that the boys' mean score was considerably higher than the girls'. The t ratio[17] (p. 324) is 2.8 which is above the 99 per cent level of significance. It appears, therefore, that when groups of this character are considered, there is a significant tendency for parents of boys to provide a better* training in the use of money than the parents of girls.

Further evidence is found for this hypothesis when the boys and girls are compared on the individual items by which parent practices were evaluated. In practically all items there was a consistent tendency for the average score of the boys to be above that of the girls.

In two items the mean score for the boys was very much higher than for the girls. These practices were: (a) parental encouragement of earning experiences and (b) parental provision for children to develop an understanding of family finances as a whole. In respect to both these practices, the number of boys exceeded the number of girls receiving positive scores.

The question may be raised as to what are the possible reasons for the boys' parents apparently providing a more valuable type of training in the use of money. Several explanations are possible.

* 'Better' to be interpreted in terms of those practices rated more valuable by the group of judges used in this study.

It could be that the boys were a more self-reliant and better-adjusted group than the girls. Upon examining the scores on the Bell adjustment and Stott Self-Reliance test, there appeared to be little difference between the scores of the boys and girls which would account for the different parental practices with these young people.

Another reason for the difference in practices might be that the parents considered it more important for the boys than the girls to be well trained in the use of money on the theory that in their future lives boys will play a larger part in the financial world than girls do. This idea was expressed by several mothers.

If parent practices and childhood experiences have an effect on the individual's later money habits and attitudes toward money management, this sex difference may be significant. In this day, the majority of girls and women have to assume considerable financial responsibility and, therefore, need as much knowledge and experience as possible.

Relationship between parent practices in teaching children the use of money and adjustment and self-reliance attained during high-school years.—It was thought that a relationship might exist between parent practices in teaching children the use of money and a child's level of adjustment or degree of self-reliance attained. Hence, coefficients of correlation (Pearsonian *r*'s) were computed for the scores on parent practices and those obtained through the use of the Bell Inventory and the Stott Self-Reliance Scale. These coefficients were all found to fall in the range $-.22$ to $+.18$. It would appear, therefore, that the kinds of experiences which these parents provide for their children in the use of money bear little relationship to adjustment or independence as measured by the two scales used. That is, there was little evidence in these data to indicate that the child given responsibility in money management was better adjusted or more independent than the one given little responsibility in the use of money.

Relationships between parent practices and later money habits of the subjects.—The data secured through the follow-up study on the actual money habits which the subjects have developed as they neared maturity and independence were evaluated quantitatively as described.

Coefficients of correlations (Pearsonian *r*'s) were determined between the scores on parent practices in training children in

the use of money and scores on later money habits. It was found that $r = +.55$ for the boys and $+.06$ for the girls. In the case of the boys, r is above the 99 per cent level of significance. Thus a definite tendency seems to be indicated in the direction of relationship between childhood experiences which parents provided or encouraged and later ability in money management. The low degree of relationship in the case of the girls may be a reflection of the 'poor' quality of experiences provided for them.

At least two factors seem to be of importance in so far as this relationship between parent practices in training boys in the use of money and the later money habits developed, is concerned. First, the parents consistently employed better methods with the boys than with the girls. The cumulative effect of this training may be indicated in the relationship between parent practices and later money habits. Second, the boys more frequently than girls were encouraged to have earning experiences and they were more frequently helped to develop an understanding of the family financial picture as a whole.

With this in mind, an effort has been made to determine the relationship between these two specific parent practices and the later habits of the subjects in the management of their own money affairs. It might be that getting out into the world, earning and accepting responsibilities that come with earning experiences, as well as being taken into the family circle to the extent that family financial planning and problems are shared between parents and children, are specific types of parental practices which are significantly related to later ability to manage one's financial resources.

Coefficients of correlation (Biserial r's) were computed, correlating the scores for the boys on each of these parent practices with the scores for the later money habits. It was found that $r_{bis} = +.35$ for the item concerned with parental encouragement of earning experiences and $r_{bis} = +.33$ for the item concerned with helping children to understand the family finances as a whole. These coefficients indicate a tendency toward possible relationship between later money habits and the parent practices mentioned above.

Further evidence for support of this hypothesis is found when the coefficient of correlation (Coefficient of Mean Square Con-

tingency*) was computed for the scores on the later money habits and the two above mentioned practices on a combined basis. In a 4 × 4 table†, $C = .48$, indicating that these two factors appear to be of importance in determining the subjects' later ability in money management.

Relationship between later money habits of subjects and adjustment and self-reliance scores.—Investigation was made of the relationship between the subjects' later money habits as represented by the quantitative score of the data secured and the measures of adjustment and self-reliance attained during the subjects' high-school period. Pearsonian *r*'s were computed.

The coefficients representing the relationship between high-school adjustment as measured by the Bell Inventory and the later money habits were as follows: (a) Home adjustment = +.40, (b) Health adjustment = +.36, (c) Social adjustment = +.24, (d) Emotional adjustment = +.37 and (e) Total Score = +.47. Except for social adjustment, these *r*'s are at or above the 99 per cent level of significance,[17] (p. 210).

These coefficients are sufficiently high to indicate a tendency toward relationship between adjustment during the high-school period and the ability to manage money in early maturity. This is not at all what would be expected in view of the fact that the data presented previously showed little evidence of relationship between parental practices in training children in the use of money and adjustment. Yet, four years later, the data indicated that those who developed desirable habits in money management tended to be the individuals who had attained satisfactory adjustment during their high-school years. It is to be noticed that this relationship was present not only for the total score representing total adjustment, but also for each area of adjustment and particularly for adjustment toward the home.

Coefficients of correlations were also computed to determine the relationship between the later money habits of the subjects and the degree of self-reliance which they attained during the high-school period in as far as self-reliance is measured by Stott's Self-Reliance Scores. Not only had the subjects rated themselves, but each mother had rated her own child.

---

* Garrett, H. E. *Statistics in Psychology and Education*, New York: Longmans, Green & Company, 1927, pp. 195–203.

† In a 4 × 4 table, *C* cannot exceed .866.

In the case of the subjects' own ratings, the coefficients were as follows: (a) Factor I, Independence = +.20, (b) Factor III, Social Resourcefulness = +.07, and (c) Factor IV, Personal Responsibility = +.27. In the case of the mothers' ratings of the subjects, the coefficients were as follows; (a) Factor I, Independence, = +.33, (b) Factor III, Social Resourcefulness = +.34, (c) Factor IV, Personal Responsibility = +.28.

The relationship between self-reliant behavior during adolescence and later money habits seemed to be insignificant for the group as a whole in so far as the subjects' ratings of themselves were concerned. There appeared to be some tendency toward a positive relationship on the basis of the mothers' ratings of the subjects.

Comparison between groups with high and low scores in later money habits of subjects.—The eight cases scoring highest and the nine cases scoring lowest in the later money habits were selected for comparison. It is recognized that the number of cases is small, yet with a homogenous sample certain tendencies may be indicated. Comparison was made on the basis of the difference between the mean scores of the high and low scoring groups in as far as the following measures are concerned: (a) parent practices in training children, (b) Bell Adjustment—Total Score, (c) Stott's Self-Reliance Scale—child's rating of himself and mother's rating of child.

This analysis indicated that in the extreme cases of the distribution, (a) parent practices in training children in the use of money, (b) the level of adjustment reached during the high-school periods, and (c) degree of self-reliance attained during high-school tend to be factors in the development of the ability to plan and manage one's income in early adulthood.

Individual case studies.—The four cases scoring highest and the four cases scoring lowest on the data secured through the follow-up study on later money habits were compared. This gave the best and poorest pictures of the use of financial resources which had been developed by these young people as they matured. These cases lend support to the results obtained through the analysis of the data on a group basis. In addition, it was indicated that: (a) the parents of the high scoring cases but not of the low scoring cases considered learning to save for future needs an important experience for children, (b) the subjects who were good managers and careful spenders as they grew into

adulthood were also good managers and careful spenders during their high-school years, while those who are having difficulty as adults in management of incomes presented the same problem during high school, (c) unsatisfactory family relationships prevailed in the low scoring groups while in the high scoring cases there seemed to be more companionship between parents and children and between the parents themselves, and (d) a more loosely woven financial pattern in the areas of budgeting and parental division of responsibility for expenditures tended to be present in the low scoring cases as compared to the high scoring cases.

SUMMARY AND CONCLUSION

The results of this investigation seem to indicate that with these young people,

a) the boys were provided with experiences which are more valuable* in training children in the use of money than the girls. While this difference held consistently in practically all of the parent practices studied, it was particularly outstanding in parental encouragement of earning experiences and in discussing family financial status, problems, expenses and goals with the children.

b) The kinds of experiences which the parents provided for their children in training them in the use of money bore little relationship to the adjustment or degree of independence shown in other types of behavior during the high-school years.

c) Parent practices in training children in the use of money tended to be positively related to later ability to utilize financial resources in early adulthood. This tendency showed up particularly in the case of the boys and in the comparative study of those individuals who displayed the 'best' and 'poorest' habits in money management as they neared maturity. It was also indicated by the correlation between later money habits and (a) parental practice of encouraging earning experiences and (b) the parental practice of discussing family financial problems and expenses with high-school-age children.

d) The money habits displayed at the beginning of adulthood tended to bear a relationship toward the type of adjustment attained during the high-school period. The conclusion that

* According to a group of experts in child development and parent education who rated parent practices in terms of value in the training of high-school children in the use of money.

'good' adjustment (especially satisfactory emotional and home adjustment) and satisfactory family relationships during childhood appear to be of some significance in the development of the ability to plan and manage one's income in early adulthood seems to be warranted.

e) The relationship between adult habits of managing income and the degree of self-reliance attained during childhood was less clear than in the case of adjustment. There were some facts which indicate that the attainment of a high degree of self-reliance in childhood may be of some importance in the development of the ability to plan and manage one's income in early adulthood. These facts stood out more clearly in the comparative study of those individuals scoring high and low in later money habits and in the mothers' ratings of their children than in the children's ratings of themselves.

These results have implications of importance for parents and teachers. The relationship demonstrated between childhood experiences in the use of money and later ability in money management emphasizes the importance of parents planning as valuable experiences as possible for children. It appears to be especially important to provide for earning experiences and to help high-school children become acquainted with family finances as a whole. Through discussion of family resources, expenses, plans and goals, adolescents may develop a better understanding of what can be afforded. It may also give them the opportunity to understand finances on a broader basis than that of their own small amount of money. In addition, it may assist in building a more closely knit family unit.

The results of this study also provide a clear demonstration of the relationship between good adjustment during high-school and the ability to manage money in early adulthood. This is one more bit of evidence indicating the inter-relationship of all aspects of living. Companionship between parents and children, sharing home responsibilities and privileges, helping children to feel that they really belong to the family unit and are partly responsible for its maintenance, play a highly significant rôle here.

BIBLIOGRAPHY

1) Andrews, B. R., "Family Relationships and Family Finances," *Child Study Magazine*, 1932, 9, 261–262.

2) Anderson, J. E., *Happy Childhood*, Appleton-Century, New York, 1938.

3) Arlitt, A. H., *The Child From One To Twelve*, Whittlesey House, New York, 1931.

4) Bell, H. M., *The Theory and Practice of Student Counseling*, Stanford University Press, California, 1935.

5) Brown, F. J., *Sociology of Childhood*, Prentice Hall, New York, 1939.

6) Charters, W. W., *Curriculum Construction*, Macmillan, New York, 1923.

7) Cole, L., *Adolescent Psychology*, Farrar and Rinehart, New York, 1936.

8) Dismorr, B., "Ought Children To Be Paid For Domestic Services," *Studies in Education*, Ed. by Earl Barnes, Second Series, 1902, p. 62–70.

9) Eliot, T. D., "Money and the Child's Own Standard of Living," *Journal Home Economics*, 1932, 24, 1–9.

10) Fisher, Dorothy Canfield, *Self-Reliance*, Holt, New York, 1929.

11) Fisher, M. H., "Are Allowances Really Practical?", *Parents*, 1937, April, 12, 24–25 ff.

12) Goodenough, F. L., and J. E. Anderson, *Experimental Child Study*, Century, New York, 1931.

13) Gruenberg, B. C., "Pennies From Heaven," *Child Study*, 1937. 14, 163–165.

14) Gruenberg, S. M., and B. C. Gruenberg, *Parents, Children and Money*, Viking Press, New York, 1933.

15) Gruenberg, S. M., "The Child's Own Money," *Child Study*, 1932, 9, 254–258.

16) Gruenberg, S. M., "Youth's Money Problems," *National Parent Teacher*, 1939, 33, April, p. 27–28.

17) Guilford, J. P. *Fundamental Statistics in Psychology and Education*, McGraw Hill, New York, 1942.

18) Hanson, Rose L., "A Study of Children's Use of Money," Univ. of Iowa Studies, Studies in Child Welfare, *Researches in Parent Education* 2, 1934, 8, 221–247.

19) Hohman, L. B., *As The Twig Is Bent*, Macmillan, New York, 1940.

20) Hoyt, E., *Consumption In Our Society*, McGraw-Hill, New York, 1938.

21) Kelley, T. L., *Statistical Methods*, Macmillan, New York, 1924.

22) Kirkpatrick, E. A., *The Use of Money*, Bobbs-Merril, Indianapolis, 1915.

23) Kohler, Anna, "Children's Sense of Money," *Studies in Education*, Ed. by Earl Barnes, 1897, 1, No. 9, 323–331.

24) Lindquist, Ruth, *The Family In the Present Social Order*, Univ. of North Carolina Press, Chapell Hill, 1931.

25) Lynd, R. S., and H. M. Lynd, *Middletown*, Harcourt, Brace and Co., New York, 1929.

26) McGinnis, Esther, "The School Child and Money," *National Parent Teacher*, 1935, Jan. 29, 20–22.

27) McLin, A. E., "Learning To Use Money," *Childhood Education*, 1940, 16, 299–303.

28) Meltzer, H., *Children's Social Concepts: A Study of Their Nature and Development*, Contributions to Education, Teachers College, Columbia University, No. 192, 1925.

29) Morgan, Winona, L., *The Family Meets the Depression*, Inst. Child Welf, Monog, No. 19, Univ, Minn. Press, Minneapolis, 1939.

30) Ojemann, Ralph H., *What Money Means To The Child*, Child Welf. Pamphlets No. 25, Univ. Of Iowa, 1933.

31) Owens, D. F., *Controlling Your Personal Finances*, Whittlesley House, New York, 1937.

32) Pace, C. Robert, "Problems in Family Relationships Among Young Married Adults," *Living*, 1940, II (No. 2, Spring,) 42–43 ff.

33) Powdermaker, F., "Setting A Price on Personality," *Child Study*, 1932, 9, 258–260.

34) Robbins, Mary, "Learning To Spend and Save," *Parents*, 1935, Dec. 10, 26 ff.

35) Sayles, M. B., *The Problem Child At Home*, Commonwealth Fund, New York, 1928.

36) Sowers, A., "Parent-Child Relationships from the Child's Point of View," *Journal Experimental Education*, 1937, 6, 205–231.

37) Stott, L., "An Analytical Study of Self-Reliance," *Journal Psychology*, 1938, 5, 107–118.

38) Symonds, P. M., "Economic Problems and Interests of Adolescents," *School Review*, 1940, 48, 97–107.

39) University of California, *Quantity and Cost Budgets For Four Income Levels*, Heller Commission for Research in Social Economics, 1939.

40) U. S. Dep't. of Agriculture, *Consumer's Guide*, 1939, 5, No. 16, P. 2, 9, Public. of Agri., Adjust. Administ., Consumers' Council Division.

41) U. S. Dep't of Labor, *Family Income and Expenditure in Chicago*, 1935–36, Bull. No. 642, Vol. II.

42) U. S. Dep't of Labor, *Family Income and Expenditures in Selected New England Cities*, 1935–36, Bull. No. 645, Vol. II.

43) White House Conference on Child Health and Protection, *The Adolescent in the Family*, E. W. Burgess, Chairman, Appleton-Century, New York, 1934.

44) White House Conference on Child Health and Protection, *The Young Child In the Home*, J. E. Anderson, Chairman, Appleton Century, New York, 1936.

45) White House Conference on Child Health and Protection, *The Home and the Child*, Martha Van Rensselaer, Chairman, Century, New York, 1931.

AN ANALYTICAL APPROACH TO THE STUDY OF READING SKILLS

WILLIAM E. HALL

Eastern Washington College of Education

AND

FRANCIS P. ROBINSON

Ohio State University

If education is effectively to teach functional reading skills, it seems important that research discover their nature and develop tests for their measurement. The evidence indicates that present-day tests (and teaching) fall far short because they emphasize simple comprehension of the single reading area of prose, and because the test elements are based on a logical analysis rather than on an experimental analysis of the actual process of reading in individuals.

During the past twenty years, the approach to constructing reading tests has tended to follow a three step evolution. The first approach was a subjective or armchair analysis of the jobs in reading for which logically consistent tests were constructed. For example, the divisions and the validity of the total score of the earliest edition of the Iowa Silent Reading Test[1] were based on the subjective analysis of reading made by Horn and McBroom in 1924.[2]

As a second step an intercorrelation technique was used to test success in making independent measures of aspects of reading, i.e., the construction of Gates Grades III to VIII Reading Test.[3] Typical of this approach is the unpublished finding of the present writers that Gates 'entirely different' reading skills correlate as

---

[1] Green, A. H., Jorgenson, A. M., Kelley, V. H., *Iowa Silent Reading Test Manual*, Revised. World Book Company, 1927, p. 3.

[2] Horn, Ernest, and McBroom, Maude, "A Survey of the Course of Study in Reading." *University of Iowa Extension Bulletin* 99, February 1, 1924.

[3] Gates, Arthur I., and Van Alstyne, Dorothy, "General and Specific Effects of Training in Reading with Observations on Experimental Technique." *Teachers College Record.* March, 1924, pp. 98–123.

Gates, Arthur I., "A Series for the Measurement and Diagnosis of Reading Ability in Grades III to VIII." *Teachers College Record*, 2, September, 1927, p. 28.

highly with each other as do typical reading tests by different authors which purport to measure the 'same' skills.

Factor analysis represents the latest step to determine what independent reading skills actually exist in individuals and what tests seem to describe them best. This approach has shown that the method of armchair analysis and labelling of reading skills is inadequate. One of the most thorough of these studies was carried out by Langsam using eight well-known tests.[4] Her results revealed five factors, i.e., verbal, perceptual, word, number, and relationship. Analysis of the tests for their loadings in each of these factors showed that almost all of them had overlapping weights in the various factors to such an extent that it was difficult to state diagnostically what a given test did measure. Gans[5] study showed that present tests tend to emphasize reading to comprehend and believe and not reading to interpret and apply. Pankaskie's[6] unpublished factor analysis of the Iowa Silent Reading Test, Iowa English Training Examination, and Iowa Mathematics Aptitude Test showed three factors, i.e., speed of comprehension, vocabulary, and ability to find answers to questions. The subtests of the Iowa Silent Reading Test did not turn out to be 'pure' measures of these factors. Another unpublished supplementary study by Pankaskie gives an intercorrelation table on Nelson-Denny Reading Comprehension and Vocabulary Tests and the Minesota Speed of Reading Test with Thurstone's seven Primary Abilities Tests. It was found that all three of these reading tests correlated significantly only with Thurstone's verbal test.

It is evident from these preliminary factor analyses that several relatively independent factors or skills are present in the reading of present-day tests. Characteristic among these are factors of 'rate of reading,' 'verbal facility' or 'vocabulary,' and 'comprehension.' It is also evident that present-day tests and their subtests do not provide 'pure' measures of these factors. Further work is needed to build better tests of the factors thus far isolated.

---

[4] Langsam, Rosalind S., "A Factorial Analysis of Reading Ability." *The Journal of Experimental Education*, 10, 1941, pp. 57–63.

[5] Gans, Roma, *Critical Reading Comprehension in Intermediate Grades*, Bureau of Publications, Teachers College, Columbia University, 1940.

[6] Pankaskie, M., "Some Factors in Reading Achievement at the College Level," *Psychological Bulletin*, 1940, 37, 573, Abstract.

Inspection of the tests used also shows that modifications and additions to the testing battery might permit the procedure of factor analysis to find other factors or skills. That is, the selections used are primarily prose and so do not permit the analysis to show factors related to non-prose reading; the test scores are products of speed and accuracy rather than treated as separate scores; and the tests mostly consist of sets of short selections with questions, whose answers can be 'looked up,' instead of long continuous selections as normally met in reading with a following quiz based on memory.

The present study makes use of previous experimental findings and attempts to answer the criticisms made above. Tests have been selected or constructed to measure factors thus far isolated. Many other tests have been added to get at additional aspects of reading and to make the determination of factors more reliable. It is felt that this extensive series of tests will, therefore, permit a factor analysis to demonstrate (1) what skills are involved in reading different prose and non-prose materials and (2) what the component factors are in such skills. This information should provide a basis for better diagnosis through indicating purer tests of different reading skills and should provide useful orientation for guiding reading instruction.

METHOD

One hundred students out of a total of one hundred sixty-five in the freshman English class at Eastern Washington College of Education acted as subjects. All were freshmen save for five sophomores who did not differ significantly in point hour ratio and age, and so were included in the final sample. The ten hours of testing for each individual led inevitably to incomplete data due to (1) some students dropping out of school, (2) others incorrectly taking the tests, i.e., double turning of pages, (3) absences at testing time, and (4) misplacing of tests. The papers for sixty-five such incomplete individuals were omitted, leaving one hundred with complete data.

Both standardized tests and ones specially constructed for this study were used. Among the published reading tests were Robinson and Hall's[7] three nonfiction tests in geology, history,

[7] Robinson, Francis P. and Hall, Prudence: *Art, Geology and History Reading Tests*, Ohio State University Press, 1940.

and art (providing separate measures of rate and comprehension accuracy over long selections); Pressey's Dictionary Test[8] (A and B forms were given and their scores combined as one test); and a test on table reading from Robinson's, *Diagnostic and Remedial Techniques for Effective Study.*[9]

In the area of charts, diagrams, and maps, adequate tests could not be found and so were constructed. The procedure was as follows: A thorough survey of texts used by students showed the type of tests needed. In every case except house plans, the illustrations used were chosen from either college textbooks or references. These tests were gone over critically by independent editors and tried out tentatively with small classes. New editions were made when these tryouts indicated such procedure would be advantageous. In some cases still further editions and tryouts were made.

The chart tests used three types of presentation (bar, line, and figure) for three subject fields (physical science, commerce, and human relations). These areas matched those tested by the Michigan Vocabulary Test and the Robinson-Hall Reading Tests. In all, eighteen charts with ten questions each were constructed to resemble typical tasks in different subject fields. These charts were scored for different forms of presentation and for different types of subject-matter; each such test had sixty questions and correlated with the other chart tests about $r = .65$. This represents a minimum reliability, since two tests in one subject-matter field or with one form of presentation would correlate higher than would two tests with different subject-matters or forms of presentation.

Three map tests were made: a flat surface map of Alaska, an ocean current map, and a street map of Denver and suburbs. Since there were twenty questions about each map, the total test contained sixty questions. Diagram tests were of four types: house plans, machines, wiring circuits, and dress patterns. The three house plans had a total of thirty questions. The machine test contained diagrams of a simple lever, a vice, and a diesel engine (ten questions each, thirty questions in all). Two wiring

[8] Pressey, S. L., *Special Reading Test Forms A and B*, Ohio State Department of Education, 1935.

[9] Robinson, Francis P., *Diagnostic and Remedial Techniques for Effective Study*, Harper Bros., 1941, pp. 137–8.

circuits were used (radio and telephone) with ten questions each. Finally, three dress patterns of common type were selected to offset the seemingly masculine loadings of the other diagrams. The pattern test had a total of thirty questions. Scoring of these tests showed that the girls made higher scores on the patterns; but there was no sex difference on the other tests.

No time limit was set on any of these tests; hence each student was measured on all items; the scores represent power rather than speed. Although these four tests (house plans, machines, circuits, and patterns) are but different types of diagrams, it was considered wise in computing the correlations to treat the scores separately.

The following items were also used as comparative or anchor points, i.e., three parts of Thurstone's Chicago Primary Abilities Tests[10] (verbal, spatial and inductive), five fields in the Michigan Vocabulary Tests[11] (human relations, commerce, government, physical science and biological science), and the student's point hour ratio.

In administering the tests, every attempt was made to secure the best possible rapport and motivation from the subjects. It was decided that the best student motivation would be secured if the testing were made as a requirement of freshman English. It was announced that the students would be given the scores if they so desired and that, if it were possible, interpretation of their test scores would be made. The total testing time was about ten hours, although any single testing period rarely exceeded an hour. The first two hours of testing was randomized for a given group and was so varied at other testing times as to eliminate test order as a factor.

RESULTS AND DISCUSSION

A factor analysis requires three steps: (1) the intercorrelation of the test variables, (2) a factoring of the correlational matrix, and (3) the rotation of axes. The twenty-five variables when intercorrelated provided three hundred correlation coefficients

[10] Thurstone, L. L., and Thurstone, Thelma G., *The Chicago Tests of Primary Mental Abilities*, The American Council on Education, 744 Jackson Place, Washington, D. C.

[11] Greene, Edward, *Michigan Vocabulary Profile Test*, World Book Co., 1944.

TABLE I.—INTERCORRELATIONS OF TWENTY-FIVE TEST VARIABLES

N = 100

Test	No.	1	2	3	4	5	6	7	8	9	10	11	12
Grade Point	1												
Machines	2	.238											
Commerce Charts	3	.391	.472										
Hum. Rel. Charts	4	.513	.386	.575									
House Plans	5	.403	.467	.547	.715								
Phys. Sci. Charts	6	.482	.559	.643	.646	.552							
Maps	7	.437	.524	.612	.595	.532	.725						
Circuits	8	.263	.323	.369	.369	.438	.398	.363					
Patterns	9	.448	.697	.104	.245	.301	.159	.203	−.018				
Speed, Geol	10	.212	.662	.099	.107	.042	.118	.142	.035	−.142			
Comprehension, Geol.	11	.408	.408	.346	.423	.319	.522	.296	.258	.257	−.285		
Speed, History	12	.283	.067	.264	.263	.145	.226	.239	.131	−.081	.705	−.028	
Comprehension, Hist.	13	.352	.395	.374	.286	.320	.419	.328	.290	.068	−.104	.550	.094
Speed, Art	14	.265	−.035	.180	.237	.060	.187	.124	.064	.041	.733	−.106	.803
Comprehension, Art.	15	.322	.150	.323	.313	.218	.359	.235	.280	.256	−.198	.431	−.065
Dictionary	16	.503	.189	.557	.540	.465	.467	.492	.219	.588	.036	.442	.054
Induction	17	.555	.223	.409	.525	.526	.502	.494	.195	.369	.204	.267	.296
Spatial	18	.174	.263	.190	.265	.214	.252	.274	.205	−.019	.019	.093	.027
Verbal	19	.469	.365	.513	.529	.372	.500	.462	.266	.199	−.116	.328	.447
Human Relations Voc	20	.359	.212	.261	.344	.234	.400	.341	.112	.302	.070	.279	.197
Commerce Voc	21	.331	.357	.289	.372	.280	.370	.324	.293	.165	−.010	.345	.205
Government Voc	22	.318	.299	.337	.374	.225	.402	.346	.230	.026	.161	.310	.339
Phys. Sci. Voc	23	.398	.373	.198	.204	.193	.456	.273	.250	.011	.004	.363	.173
Biol. Sci. Voc	24	.465	.087	.319	.220	.329	.266	.321	.088	.447	.064	.262	.215
Tables	25	.301	.268	.213	.338	.374	.426	.352	.187	.231	.082	.343	.118

TABLE I.—*Continued*

Test	No.	13	14	15	16	17	18	19	20	21	22	23	24	25
Comprehension, Hist.	13													
Speed, Art.	14	−.020												
Comprehension, Art.	15	.520	−.166											
Dictionary	16	.123	.033	.282										
Induction	17	.168	.269	.210	.688									
Spatial	18	.046	.059	.024	.090	.210								
Verbal	19	.392	.394	.337	.445	.395	.193							
Human Relations Voc.	20	.407	.238	.337	.212	.301	.130	.414						
Commerce Voc.	21	.482	.199	.401	.087	.217	.199	.563	.392					
Government Voc.	22	.504	.244	.325	.232	.267	.076	.461	.423	.454				
Phys. Sci. Voc.	23	.303	.054	.230	.003	.145	.210	.311	.230	.261	.448			
Biol. Sci. Voc.	24	.211	.084	.198	.203	.312	.018	.375	.279	.197	.200	.305		
Tables	25	.506	.096	.249	.208	.219	.109	.324	.191	.322	.438	.341	.288	

found by the Pearson product method. An independent check of all computations was made to insure accuracy. These correlations are presented in Table I.

The correlations vary from .80 to −.285. The average correlation of the reading tests (omitting point hour ratio and Thurstone's Verbal, Inductive and Spatial Tests) is .28. This average indicates a small amount of communality of skill or ability and suggests that there must be considerable independence of skills. It is obvious that the nature of this communality and independence is not discernible until after further analysis.

The second step in Thurstone's centroid method of factor analysis involves the estimation of communalities and the computing of factor loadings and residuals. The result of this procedure revealed six factors with their loadings.

The third step in the analysis is the rotation of the factors into simple structure. The orthogonal reference frames were rotated with regard to the configuration of the twenty-five test vectors in six dimensions, until the coördinate axes became most meaningful. This process demands some trial and error application of the formula, and the results of those applications determine which method provides the best fit. In all, twenty rotations were made. The factor loadings and communalities after rotation are given in Table II; they vary from −.66 to .83.

Certain principles are commonly accepted by factor analysts in interpreting factor loadings of tests: (1) A test with a loading of .400 and above has real significance in meaning for that factor, but a test with a loading of .100 or less has zero or negligible significance. (2) To be described a factor must be over determined; namely, it must have a weighting of .400 or more in at least three tests. (3) The absence of loading in tests may aid interpretation as well as the presence of loading in tests.

Factor No. 1 may be identified as a study attitude. It will be noted in Table II that the highest loadings are in tests on comprehension accuracy (11, 13, 15). These tests ask the student to read a selection in order to answer a quiz; his score is the per cent right of questions over material read. The material read is not particularly difficult. The student selects a level of care in reading which he believes adequate to succeed on a typical quiz; the per cent right indicates the level of comprehension accuracy at which he was reading. The absence or negligible weight of

TABLE II.—FACTOR LOADINGS OF TWENTY-FIVE VARIABLES AFTER ROTATION

Test		Factors						
		1	2	3	4	5	6	h^2
1	Point Hour	.1916	.3590	.3440	−.0841	.4980	.2035	.5804
2	Machines	.0880	−.0325	.8352	−.3147	.0446	.0722	.8126
3	Commerce Chart	.1791	.1012	.4522	.1826	.3179	.4473	.5840
4	Human Relations Chart	.1441	.1015	.5154	.2012	.3866	.4506	.6897
5	House Plans	.3354	.2387	.4907	.0123	.0202	.5333	.6953
6	Science Chart	.3844	.0644	.5385	.0288	.2931	.4501	.7312
7	Maps	.2429	.1653	.5305	.0358	.2261	.4698	.6408
8	Circuits	.2388	−.1146	.2901	−.0490	.1861	.3729	.3304
9	Patterns	−.1551	.2308	.2191	−.7775	.4600	.2369	.9976
10	Geology Speed	−.1204	.0216	.0206	.2727	.7919	−.4743	.9419
11	Geology Comp	.5858	−.1792	.2414	−.0524	.3581	.0105	.5646
12	History Speed	.0910	.3680	.0870	−.3146	.6710	−.3357	.8132
13	History Comp	.5604	−.0800	.5561	.0478	−.0507	−.0306	.6354
14	Art Speed	.0851	.4705	−.0224	−.3742	.6422	−.3787	.9249
15	Art Comprehension	.4859	−.1726	.1982	.1808	.3355	−.0600	.4541
16	Dictionary	.3078	.4766	.3015	.5432	.1213	.1307	.7802
17	Induction	.3438	.5118	.3401	.2887	.2753	−.0257	.6556
18	Spatial	.1165	−.0208	.1220	−.0045	.1454	.2579	.1165
19	Verbal	.0815	.1451	.7431	.0599	.2775	.1658	.6880
20	Human Relations Voc	−.0144	.1820	.4668	.0232	.1954	.2663	.3608
21	Commerce Voc	.2355	.1018	.6365	.2072	.0627	−.0814	.5141
22	Government Voc	.4042	.2452	.5515	.1519	−.0813	.0836	.5644
23	Physical Science Voc	.3004	−.0849	.3824	.3115	.1359	.0208	.3595
24	Biological Science Voc	.2508	.3420	.2284	−.1047	.2774	.2729	.3946
25	Tables	.2879	−.0658	.4234	.2213	.2216	.0311	.3656
							Mean h^2	.6078

the rate tests (10, 12, 14) in this factor is of interest especially in view of the fact that too frequently reading tests combine rate and comprehension into one score.

The Greene Vocabulary Test on government has a significant weighting here, but the others in this series do not (.01, .23, .30). The test of government vocabulary was the easiest of the five tests; possibly for this reason attitude toward the test might be important.

Factor No. 2 seems best identified as an Inductive Factor. The tests with the highest weights are Thurstone's Inductive Test, .5118, Pressey's Dictionary Test, .4766, and Robinson-Hall's Rate Test in Art, .4705. Another rate test (history 12) has a loading of .3608 and point hour ratio (1) has a loading of .3590. These last two are included, even though they do not reach the standard of significant loadings, because they seem to help describe the factor.

It seems reasonable to label factor No. 2, the Inductive Factor, because Thurstone's Induction Test (developed from other factor analysis studies to measure one independent factor) had the highest loading. Thurstone reports that the inductive factor appears most significantly in tests in which the subject is required to find a rule that binds a number of items together and then use the rule in classifying or predicting other items. Pressey's Dictionary Tests are definitely of this nature. The text used in Robinson-Hall's Art Test marshalls facts that lead up to a generalization, i.e., inductive writing. It should be noted that the Art Rate Test is not simply a measure of speed, but of the rate of comprehending ideas presented inductively (at a comprehension level acceptable to the reader). On the other hand, the Geology Rate Test in this same series has a zero weight in this factor; the speed of comprehending this dry factual presentation of isolated details in each geological epoch seems to be a different factor (see factor #5).

Although point hour ratio does not quite have a significant loading, it will be noted in Table I that the inductive test correlated the highest of all twenty-four tests with point hour ratio ($r = .55$).

The tests without loadings also help in this description. It will be noted that tests of comprehensive accuracy (11, 13, 15) that were highly loaded in factor #1 have zero loadings in this

inductive factor. Again it is shown that factor #1 is more an attitude toward care in reading rather than ability to comprehend. The Machine and Physical Science Chart Tests also have a zero loading; in fact all the chart tests have negligible loadings. The vocabulary tests generally have low or zero weights, although biological science shows some strength (.3420).

Factor #3 seems best identified as a Verbal Factor. Without question this is the most strongly determined of all factors, being significantly loaded in thirteen tests. It also accounts for the greatest amount of variance (average h^2). It seems without doubt to be easily described as the verbal factor; it has high loading on Thurstone's Verbal Test (19) and significant loadings in four of Greene's Vocabulary Tests (20, 21, 22, 23). Test 2, 3, 4, 5, 6, and 7—although chart tests—are all significantly loaded, probably because they all have verbal questions and labels. Tests proved in other studies to be loaded with such a verbal factor are: reading tests, verbal analogies, synonyms, opposites, vocabulary, grammar and verbal associations.

It will be recalled that comprehension accuracy tests (11, 13, and 15) were highly loaded in factor #1, but the rate tests (10, 12, and 14) and Thurstone's Verbal Test (19) had a zero rating or a very low loading. In factor #2, the rate tests had high loadings, but comprehension accuracy and Thurstone's Verbal Test had insignificant loadings. And in factor #3, Thurstone's Verbal Test has a heavy loading and the comprehension accuracy and rate of reading tests have zero or low loadings except for History Comprehension Accuracy (.5561). This surely indicates that comprehension accuracy, rate of reading, and verbal knowledge should not be called simply aspects of the same skill and lumped together in one test score as is frequently done in present-day tests. They represent three different factors.

Factor #4 introduces some results which do not seem to fit orthogonal projection; it seems worth while to postpone this problem for the moment until factors #5 and #6 have been discussed.

Factor #5 may be identified as another rate factor. It definitely has high loadings in tests 10, 12, and 14, which measure reading speed. But it seems clearly different than the inductive factor #2, which had its highest loading in the art rate test. Geology rate was practically absent in factor #2 but is prominent in factor #5. In fact, the order of loading in factor #2 was, art

rate, history rate, and geology rate; in factor #5 this order is reversed. The style of writing in the art and the geology texts are significantly different. The geology text is written as a condensed summary of the many elements (relationships not indicated) in each geological epoch and the art in an inductive style as before mentioned. These text differences provide us with a cue for describing factor #5, although it should be noted that all three rate tests have high loadings on this factor. This factor might be called the non-inductive-prose rate skill to distinguish it from factor #2.[12] The meaning of the point hour ratio having a significant weighting in this factor is not clear.

Factor #6 may be identified as a chart reading skill, although it is apparent that all charts do not take a single skill. Of the eight chart tests, five have significant loadings and a sixth approaches significance; the two exceptions are machines and patterns which did not behave in the same way as the other chart tests. It will be noticed that the test on wiring circuits and Thurstone's Spatial Tests both have their highest loadings in this factor (.3729 and .2597). This might suggest some small relationship to a spatial factor. Factor #6 has next to the most tests with zero loadings; the usual prose reading tests do not touch this factor at all.

If it were not for the three rate tests, factor #6 would satisfy all criteria, i.e. maximized zero loadings, as many positive loadings as possible and no negative loadings. On the other hand, such definite clusterings of positive, zero and negative loadings surely indicates structure of some type. While not quite according to the rules, the writers cannot help but see an antithesis in the positive cluster (chart tests) and the negative cluster (prose tests). Certainly the positive cluster is independent of zero loadings; probably the negative cluster is also (factor #5 at least shows rate to be an independent factor). But do the results here mean that students who had chart skill as measured by

[12] Factor #5's relationship to the speed factor described by Thurstone and Langsam likewise is not clear. Their factor is described as the "facility of perceiving detail imbedded in irrelevant detail" and as "a function that appears to be a facility in perceiving detail." Langsam found the Iowa Silent Reading Rate Tests, Word Meaning, Michigan Speed of Reading Test, A.C.E. Verbal Analogies Test and Thurstone's Identical Forms test to be significantly loaded in her speed factor. None of these tests are included in the present study.

factor #6 were deficient in prose rate skill? Is this a question of visual-mindedness versus verbal-mindedness? Or is it that some students' interest and training in mechanical things preclude success along verbal lines? The answer cannot be given, but questions raised by this unusual statistical structure should not be ignored.

Factor #6 is definitely not a rate of reading prose; it is also definitely independent of verbal factors. With the negative rate factor, the absence of verbal and inductive loadings, and the high chart loading, one may with some confidence interpret it as a skill, probably visual- or mechanical-mindedness, entering into chart reading.

Factor #4 is not here identifiable. That is, this factor is insufficiently determined with only two tests (9 and 16) having a significant loading; one of these is positive and the other negative—a situation which also makes interpretation in terms of orthogonal projection impossible. Since test 9 contains dress patterns, this factor may be related to sex, but it cannot be said that the dictionary test, which was positively loaded, is a male speciality. It seems best merely to say that factor #4 is a very cloudy factor and stands in the need of further investigation.

CONCLUSIONS AND IMPLICATIONS

Factor analysis has been widely accepted as a means of exploring uncharted domains where basic and fruitful concepts are essentially lacking. The findings of this study show the existence of the following factors in the types of reading accuracy situations studied:

1) Attitude of Comprehension Accuracy
2) Rate of Inductive Reading
3) Verbal or Word Meaning
4) Not Identifiable
5) Rate for Unrelated Facts
6) Chart-reading Skill

One of the clearest implications of this study seems to be in the fact that there are different reading skills for prose and for non-prose materials. Each of these skills is itself a composite of different subskills, abilities and attitudes. Four of the five factors found seem to be coördinated into 'prose reading': one cannot read prose without using at the same time the factors of

(1) comprehension accuracy, (2) inductive rate and (3) word meaning; or, with a different type of prose, using at the same time (1) comprehension accuracy, (3) word meaning, and (5) non-inductive rate. Two factors are coördinated as independent components into 'chart' reading. One cannot read charts or graphs without combining the factors of (3) verbal skill and (6) chart skill. In pattern reading it was found that a different pattern of factors acted as components. These components were factor #5 and the muddy factor #4 (whatever it represents). Reading tests should not stick, therefore, to prose and other strictly verbal materials.

It is usual for comprehension accuracy, rate, and verbal factors to be thrown together in a test to make one score. Thus, when time limit tests are scored as the number of questions answered correctly, this is a product of the number of items covered (rate), attitude toward accuracy, and ability to comprehend (verbal ability). Independent scores should be provided for these three variables.

The findings have further implications for testing: (a) It seems evident that the tests on the market are in many cases mis-labelled; the traits they claim to measure do not show up as independent skills. (b) There appears to be ample evidence that the range of tests used in measuring reading skills should be greatly expanded to include such different types of reading as prose, charts, maps, indices, blue prints and tables. Such tests as the Iowa Every-pupil Test and the Chicago Reading Tests are to be commended for their start in this direction. (c) As far as this study was concerned, differences in subject-matter did not play as dominant a röle in differentiating reading skills as did other factors.

The first step in future research would seem to be an attempt to purify the tests which have been identified here as independent components of larger reading activities. The second step would appear to be further research as to skills that have as yet not been identified in reading, such materials as blue prints, geodetic maps and aerial maps.

EQUATING HIGH-SCHOOL INTELLIGENCE QUOTIENTS WITH COLLEGE APTITUDE TEST SCORES

EDMUND G. WEBER

Senior Class Advisor, Central High School, Minneapolis, Minnesota

Although some form of intelligence tests have been given in the Minneapolis schools for many years, educators in general and counselors in particular feel that there is a need for a better understanding of the meaning of intelligence quotients in terms of college aptitude test scores. At the present time, the intelligence tests are usually administered to pupils at various intervals up to and including the tenth grade. The college aptitude test is always given sometime during the twelfth grade. Hence, there is a span of years in which, because of the lack of full knowledge concerning the meaning of the intelligence quotient in terms of college aptitude test scores, counselors are not able to take full advantage of all of their counseling aids.

Since secondary schools, colleges, and other educational institutions place considerable stress upon college aptitude test scores, it seems proper to suggest that a table or graph comparing equivalent intelligence quotients and college aptitude test scores would fill this need. To be more specific, by the use of such an aid, counselors and others interested could predict the probable college aptitude test score from any given intelligence quotient. What is more important, this prediction could be made at any time after the intelligence test had been taken. It can easily be seen what a motivational stimulus this could be in the hands of competent and well qualified counselors. This report is an attempt to fill this need.

THE PROBLEM

The purpose of this paper is to present the method by which the intelligence quotients of graduates of Minneapolis public high schools were equated with the scores made by them on the American Council on Education Psychological Examination.

THE PROCEDURE

The first step in equating the test scores was to locate comparable records for at least one thousand Minneapolis high-

school graduates who later attended the University of Minnesota. Further, if the findings were to be used as a counseling aid to suggest future University training, all of the cases selected for study would have to be graduates who actually did attend the University. The most accessible source of data proved to be the photostatic copy of credits earned in the University by each student, which is sent to the high school from which he graduated at the end of the freshman year. It was found that three of the ten high schools had not preserved this record; hence the data are based upon the records from the seven schools which had kept the reports. These high schools were Central, Edison, Roosevelt, South, Southwest, Washburn, and West.

Since over a period of years many different mental ability tests have been used in the local schools, the next problem was to determine a single intelligence test upon which efforts at equating scores could be centered. However, this proved to be no problem, because during the past five years the guidance department has concentrated on the Otis Quick-Scoring Tests of Mental Ability. Hence, all intelligence quotients used in this report were derived from this Otis test. Further, the pupils were all given this test between their thirteenth and sixteenth years. The tests were all administered by competent school authorities. These scores were then posted on the cumulative record cards and kept on file in the school office for a period of five years after graduation.

The graduates of the classes of 1940–42 were selected because—(1) they all had the same test, (2) they were pre-war classes that had the normal proportions of sexes entering the University, and (3) there was no reason to expect to find some factor which would make these groups un-normal.

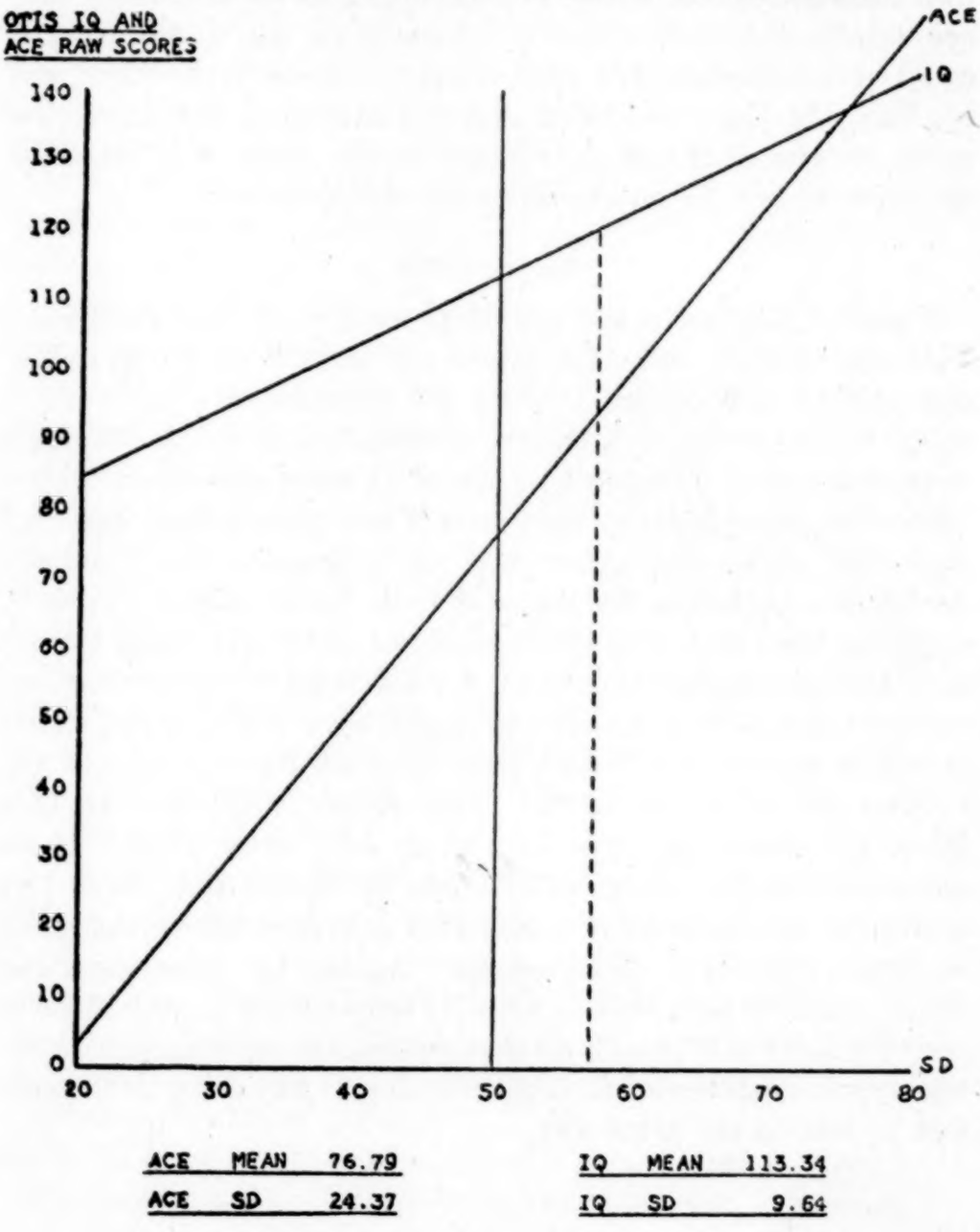
The college aptitude test scores were obtained from the American Council Psychological Tests. This test is given to all Minneapolis high school seniors by the University of Minnesota authorities during their last semester in school. It takes sixty minutes to administer this test, which is composed of the following sections: (a) completion, (b) arithmetic, (c) artificial language, (d) analogies, (e) opposites. The test papers were machine-scored in the University Testing Bureau of the University of Minnesota.

The author personally gathered all of the data used in this

report. Each graduate's record was carefully studied to be sure that only those were included who conformed to the limitations which had been set up in selecting the population for the study. It was found that 1014 graduates met these limitations. Because of the fact that the University of Minnesota does not differentiate between the sexes, no attempt has been made to do so here. The

FIGURE I

OTIS IQ AND ACE RAW SCORES OF UNIVERSITY OF MINNESOTA FRESHMEN EQUATED BY THE STANDARD DEVIATION LINEAR TECHNIQUE



ACE	MEAN	76.79	IQ	MEAN	113.34
ACE	SD	24.37	IQ	SD	9.64

combined scores of both boy and girl graduates were used in all computations.

Inasmuch as we found a relatively normal distribution of test scores, the Standard Deviation Linear Technique was chosen as the best method for equating the two variables. Therefore, we equated by transposing both Otis intelligence quotients and American Council Psychological raw scores into standard score equivalents. In making this transformation, we accepted the following assumptions: (1) that the raw score units are equal, (2) that the mean scores of the two distributions are equal, (3) that one standard deviation above the mean on one distribution is equal to one standard deviation above the mean on the other, and (4) that, therefore, one-tenth of a standard deviation above the mean on one distribution is equal to one tenth of a standard deviation above the mean on the other distribution.

THE METHOD

Figure I illustrates the graphical version of this technique. Both the Otis IQ and ACE scores are plotted on the ordinate, one unit to each square. Along the abscissa, two squares are equal to one-tenth of a standard deviation unit. A standard deviation unit of 50 represents the mean score of both variables. Therefore, a standard deviation unit of 60 represents one standard deviation above the mean, and 40 represents one standard deviation unit below the mean of both distributions. Thus in applying the chart, if a line is dropped anywhere along the IQ axis, the comparable ACE score will be found wherever that line contacts the ACE axis. For example, by consulting the chart, it will be seen that a line dropped from an IQ score of 120 will contact the ACE axis at 94. This means, then, that an Otis IQ of 120 should be equivalent to an ACE score of 94. If we continue this line downward, it can be shown that these two scores are seven-tenths of a standard deviation above the mean of their respective distributions. Again, by consulting the chart, it can be seen that an Otis IQ score of 104 is equivalent to an ACE score of 52, and both these scores are one standard deviation below the mean of their distribution. Any other ACE score can be read in the same way.

PUBLICATIONS RECEIVED

DORA A. AMES AND OTHERS. *Meet the Soviet Russians.* Cambridge, Mass.: Graduate School of Education, Harvard University, 1944, pp. 89. (paper)

EDWARD L. BERNAYS. *Take Your Place at the Peace Table.* New York: The Gerent Press (270 Madison Ave.), 1945, pp. 60. (paper)

EMMETT ALBERT BETTS AND THELMA MARSHALL BETTS. *An Index to Professional Literature on Reading and Related Topics.* New York: American Book Co., 1945, pp. 137. (paper) $1.50.

BERNARD CARP. *A Study of the Influence of Certain Personal Factors on a Speech Judgment.* New Rochelle, N.Y.: The Little Print, 1945, pp. 122. (paper) $1.75.

H. HARRISON CLARKE. *The Application of Measurement to Health and Physical Education.* New York: Prentice-Hall, Inc., 1945, pp. 415. $3.75.

GEORGE S. COUNTS. *Education and Promise of America.* New York: The Macmillan Co., 1945, pp. 157, $1.50.

ROY M. DORCUS AND G. WILSON SHAFFER. *Textbook of Abnormal Psychology.* Third Edition. Baltimore: The Williams and Wilkins Co., 1945, pp. 547. $4.00.

EDGAR H. DUNCAN, Editor. *The Humanities in Higher Education in the South.* Report of a Conference held at Vanderbilt University, July 24–29, 1944. Nashville, Tenn: Vanderbilt University, 1945, pp. 50. (paper) $0.25.

T. L. ENGLE. *Psychology: Principles and Procedures.* Yonkers, N. Y.: World Book Co., 1945, pp. 549. $2.12.

EDITH GANN. *Reading Difficulty and Personality Organization.* New York: King's Crown Press (Morningside Heights), 1945, pp. 149. (paper) $2.00.

WARREN H. GARDNER. *Left-handed Writing: Instruction Manual.* Danville, Ill.: The Interstate, 1945, pp. 28. (paper) $0.60.

LEO IVOK. *How to Prepare the Schedule for a Secondary School.* Cambridge, Mass.: Graduate School of Education, Harvard University, 1945, pp. 26. (paper)

DALLAS JOHNSON. *Veteran's Guide.* New York: Public Affairs Committee, Inc., 1945, pp. 32. (paper) $0.10.

PHILIP MURRAY. *C. I. O. Re-employment Plan.* Washington,

D. C.: C. I. O. Department of Research and Education, 1945, pp. 29. (paper) $0.15.

RUDOLF PINTNER AND ARTHUR I. GATES. *The Value of Individual Hearing Aids for Hard of Hearing Children.* Washington, D. C.: National Research Council, 1944, pp. 40. (paper)

VAHAN D. SEWNY. *The Social Theory of James Mark Baldwin.* New York: King's Crown Press (Morningside Heights), 1945, pp. 93. (paper) $1.50.

MAXINE SWEEZY. *Medical Care for Everybody.* Washington, D.C.: American Association of University Women, 1945, pp. 38. (paper) $0.15.

SISTER ROBERTINE WEIDEN. *The Effect of Checked Directed Study upon Achievement in Ninth-grade Algebra.* Baltimore: Johns Hopkins Press, 1945, pp. 85. (paper) $1.25.

ALEXANDER G. WESMAN. *A Study of Transfer of Training from High School Subjects to Intelligence.* New York: Bureau of Publications, Teachers College, Columbia University, 1945, pp. 82. $1.75.

RANYARD WEST. *Conscience and Society: A Study of the Psychological Prerequisites of Law and Order.* New York: Emerson Books, Inc. (251 W. 19th St.), 1945, pp. 269. $3.00.

EDWARD A. WIGHT. *Evaluation and Revision of the Library School Curriculum.* Nashville, Tenn.: Peabody Press, 1945, pp. 64. (paper) $1.10.

BEN D. WOOD, Director. 1944 *Fall Testing Program in Independent Schools and Supplementary Studies.* New York: Educational Records Bureau (437 W. 59th St.), 1945, pp. 49. (paper)